

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## THREE TRILLION AND COUNTING

Mapping tree density at a global scale **PAGES 170 & 201**

### ARCHAEOLOGY

#### LOST WET WORLD

Deep search for evidence of a pre-Clovis culture

**PAGE 176**

### INEQUALITY

#### CLOSING THE GENDER GAP

Why women get fairer share of social-sciences funding

**PAGE 181**

### BOOKS

#### GENES AND THE MAN

Reflections on Dawkins's latest autobiography

**PAGE 184**

**NATURE.COM/NATURE**

10 September 2015 £10

Vol. 525, No. 7568





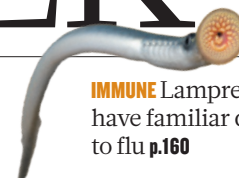
# THIS WEEK

## EDITORIALS

**QUESTIONS** Is this brilliant editorial the best ever written? **p.158**

**WORLD VIEW** Policy has its own problems to reproduce **p.159**

**IMMUNE** Lamprey fish have familiar defence to flu **p.160**



## Keep a welcome

*The plight of a record number of refugees is something the West cannot ignore. Humanitarian values should be upheld, and people fleeing war and persecution must be offered protection.*

In the refugee crisis facing Europe and the Middle East, an image can be worth a thousand articles or opinion pieces. Academics and humanitarian organizations have long battled to debunk the vicious myths and disinformation that often surround the refugee issue, and to counter often fact-free government policies — to little effect. It took a single iconic and heartbreaking image of a three-year-old Syrian boy, Aylan Kurdi, washed up drowned on a beach in Turkey for the world's conscience to wake up to the plight of refugees.

Science and other academic interests have a long tradition of offering both refuge and professional hope to displaced people. Almost every discipline has its own story of influential figures in the field who arrived with oppression and conflict snapping at their heels. This journal has long chronicled and supported such efforts. In June 1939, for example, *Nature* published a three-page editorial that concluded that if Britain relaxed its “exceedingly cautious” attitude to accepting refugees, then this would not only defend humanitarian values and academic freedom, but also “might prove in the long run to be wise and sound from the economic point of view”.

What is there to say in 2015? Worldwide, there are some 60 million refugees, up from 37.5 million a decade ago — the biggest refugee crisis since the Second World War. Yet the humanitarian response so far has been largely inadequate. The shrill rhetoric in many European Union and other wealthy nations claiming an ‘invasion’ of refugees doesn’t stand up to scrutiny. Four million refugees have fled Syria since the conflict began there in 2011, but last year, the United Kingdom accepted 4,500 Syrian refugees, or just 0.007% of the UK population. Among the more generous EU countries, Germany took in 40,000 and Sweden 34,000 — the United States took only 4,750. By contrast, over the same period, Turkey temporarily accepted 1.5 million, and Lebanon, a tiny country of just 4.5 million people, took in some 1.15 million refugees, or 26% of its population.

EU refugee law is a mess. For refugees to apply for asylum, they must first reach a territory outside their own country. But the EU, and other countries, have increasingly sought to circumvent international refugee law by introducing rules to keep refugees out and so prevent them from applying in the first place.

A pernicious 2001 EU directive, for example, erects a barrier by imposing fines and the costs of repatriating illegal immigrants on airline, train, shipping and other carriers, essentially shifting the responsibility for deciding who is a legitimate refugee and who is an illegal migrant from governments to the carrier companies. Predictably, carriers have refused to accept passengers who lack visas. This fortress-Europe mentality explains why, this year alone, more than 300,000 people have embarked on perilous crossings of the Mediterranean — with 2,600 perishing — instead of taking a commercial ferry or airliner to apply for asylum.

There is also no EU-wide asylum status, with decisions on applications left to each member state, and no mutual recognition of positive

outcomes by countries. And the seriously flawed ‘Dublin Regulation’ also obliges the EU member state in which a refugee first arrives to take the refugee’s asylum application. This has resulted in frontier countries such as Greece and Italy bearing a hugely disproportionate burden.

The rule also frustrates applicants who have a legitimate preference for a specific country, for example to join their extended family. This encourages irregular movement within the EU, and allows other mem-

ber states to forcibly return refugees to their first port of call — so turning what should be a humanitarian exercise into one of excessive coercion and criminalization.

**“The EU, and other countries, have increasingly sought to circumvent international refugee law.”**

In August, Germany’s Chancellor Angela Merkel rightly suspended her country’s adherence to the Dublin Regulation, and called for a radical, permanent EU-wide system of processing asylum applications,

with an enforced distribution of refugees throughout EU member states. Merkel last week courageously stated that Germany itself can and will cope with its inflow of refugees, an expected 800,000 this year. The proposal is vigorously opposed by some member states, in particular the Czech Republic, Poland, Hungary and Slovakia.

The public outcry following the photograph of Aylan has given the proposal new momentum, with François Hollande, the French president, last week lending his support, and also the United Nations. The EU will formally discuss the proposal on 14 September — it should be embraced as long-overdue reform.

The scientific community must also play its part. It is in everyone’s interest for refugee students and academics to be given opportunities to continue their careers, because otherwise, the Middle East and elsewhere risks losing a generation of talent. The Western academic community must boost efforts to welcome refugee academics and students. ■

## Money matters

*It is not how much people have, it is how much we know they have that stokes inequality.*

It would be so convenient if fundamental laws of nature told us how best to run a society. Governance would be a simple optimization problem, like finding the shortest route through a network; we could do without left-right political confrontation, and just solve the equations. Unfortunately, governance is not a well-posed problem. There must inevitably be balance and compromise: for example, of the rights of the individual against the overall good for society. This is what

makes politics and economics not just controversial, but interesting.

Inequality is one of the biggest items on the agendas of both of these disciplines. Few people are likely to speak in favour of inequality as such, but in stereotypical terms the political right defends wealth as a reward for hard work, whereas the left deplores a society in which, as economist Joseph Stiglitz has said of the United States, “1 percent of the people take nearly a quarter of the nation’s income”. It seems an unavoidable truth that a free-market capitalist system will create wealth inequality; to a free-market fundamentalist who sees markets as meritocratic optimizers of efficiency and resource utilization, that is not only necessary but moral. Under that philosophy, by intervening in the market in the hope of making the outcome ‘fairer’, we only throw a spanner in the works.

Yet even if one accepts some inequality as a necessary evil, there are options beyond *laissez-faire*. How, and how strenuously, governments and legislators should attempt to limit the extent of wealth inequality — crudely measured by the Gini coefficient, which quantifies the statistical dispersion of income distribution — is currently a hotly disputed matter. Should companies and banks be restricted in what they can pay their chief executives? Should taxes aim to inhibit or reduce the perpetuation of inherited wealth? Or is all this crypto-communist social engineering?

The strongest argument for such measures is not that it makes things more ‘fair’ (although meritocratic defences of free-market inequalities should surely at least demand a level playing field). Rather, it is that gross wealth inequality is socially corrosive. It polarizes attitudes, foments unrest (see, for example, the Occupy movement) and degrades trust and cooperation. At face value, a study published online this week in *Nature* supports that view — but with an added twist.

In the study, groups of volunteers played a simple economic game involving cooperation (a “public goods game”), in which they could lose or gain wealth through voluntary redistribution within social networks that started with three different levels of inequality (A. Nishi *et al.* *Nature* <http://dx.doi.org/10.1038/nature15392>; 2015). Crucially, in some games the wealth of participants was made visible to others, whereas in others it was kept hidden. For “invisible” wealth conditions, the games tended to converge on a fairly low Gini coefficient,

but “visible” wealth produced higher (and less stable) average Gini coefficients. This result was exacerbated when the initial inequality was greater. In other words, simply hiding wealth decreased the wealth disparity in otherwise identical games and networks.

Still more importantly, visible wealth reduced the overall cooperation and interconnectedness of the social network, and in fact led to lower total wealth. As the authors say: “it is not inequality per se that is so problematic, but rather visibility” of that inequality. This fits with the established idea that it is relative, not absolute, differences in wealth that compromise happiness and promote discord: we resent what our neighbours have and we don’t. What grates is not knowing that others have more than us, but seeing that difference ostentatiously displayed.

It is dangerous, however, to think that these laboratory experiments can be extrapolated into a political or moral message for the real world. They invite us to frown on bling and the champagne-drenched excesses of financiers, but we should be cautious about their implications, even (or especially) if they flatter our preconceptions. Besides, there is scope here for upsetting both ends of the political spectrum. Right-wingers might deplore an injunction to hide one’s wealth, compromising personal freedom — isn’t it up to us how we spend our money? Left-wingers might dislike the idea of being relaxed about inequality as long as it is kept out of sight — and, anyway, might that not provoke a climate of secrecy and suspicion?

For now, the results should simply inform and broaden the discussion. They show, for example, that inequality is not solely down to market mechanisms, but also responds in subtle ways to our own dispositions. Above all, the findings are a reminder, along with related behavioural experiments on the role of punishment in public-goods games, that John Maynard Keynes’s “animal spirits” are an irreducible part of what shapes a market economy. It is time to lay the idea of the rational *Homo economicus* to rest. ■

**“Inequality is not solely down to market mechanisms, but also responds in subtle ways to our own dispositions.”**

## Loaded language

*There can be more to a question than appears at first sight.*

William Burroughs, the infamous US writer and author of *Naked Lunch*, had a typically counter-culture approach to seeking knowledge: “Your mind will answer most questions if you learn to relax and wait for the answer.”

If only it were that easy for the rest of us. Instead, to ask a question is harder than it might seem. British Prime Minister David Cameron discovered this last month when the UK Electoral Commission told him to change the wording of a proposed question for the country’s referendum on membership of the European Union.

Cameron’s suggestion — “Should the United Kingdom remain a member of the European Union?” — was a classic example of what linguists call acquiescence bias. Take the Burroughs route and relax, and the answer to such a question that comes to mind more often than not is to stick with the status quo. Rejecting something is more difficult.

If that was Cameron’s intention, then his plan has been rumbled. The question will now have the extra clause at the end: “or leave the European Union?” To answer that one, citizens must now make more of a cognitive effort, and that should remove the chance for bias.

Cameron’s linguistic nudging was more subtle than most attempts to bias questions. Lawyers and politicians tend to be fans of more explicit tricks of language. There is the classic loaded question — when did you

stop beating your wife? — which presupposes guilt; and the pernicious influence of the hypothetical question. During the 2000 US election campaigns, South Carolina voters were asked: would you be more likely or less likely to vote for John McCain for president if you knew he had fathered an illegitimate black child??

Researchers have found that the way a question is phrased can alter how people remember incidents. Witnesses asked how quickly cars were travelling when they “smashed” are more likely to imagine that they saw broken glass on the ground than others told that the vehicles simply “bumped” into each other or “collided”. They were also more likely to say that the cars were travelling at higher speed.

Scientists have a particular relationship to questions. Turned into testable null hypotheses, questions are at the heart of the scientific method. Allied with proper experimental design and robust statistical analysis, they can be answered with confidence — or not.

Some answers are known before the question is asked; other questions are genuine calls for information. Some want to benefit the questioner and others to empower those who answer it. How to judge? In all areas — politics and science included — the best questions are simple and to the point. So who knows what the residents of Quebec thought when confronted with the following for their referendum on independence in 1995:

“Do you agree that Quebec should become sovereign, after having made a formal offer to Canada for a new economic and political partnership, within the scope of the bill respecting the future of Quebec and of the agreement signed on June 12, 1995?”

The ‘No’ vote won with 50.6%. ‘Don’t know’s were not recorded. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)



MICHAEL TEMCHINE



## Reproducibility will not cure what ails science

A bill to make data for environmental regulation more transparent reveals the fuzzy boundary between science and ideology, argues **Daniel Sarewitz**.

Leaders of the scientific community, nudged by the media (including *Nature*), are acknowledging that a culture of science focused on rewarding eye-catching and positive findings may have resulted in major bodies of knowledge that cannot be reproduced.

Private-sector, academic and non-profit groups are leading multiple efforts to replicate selected published findings, and so far the results do not make happy reading. Several high-profile endeavours have been unable to reproduce the large majority of peer-reviewed studies that they examined. Meanwhile, the US National Academies is preparing to publish a high-profile report on scientific integrity that will flag irreproducibility as a key concern for the research enterprise.

As the spotlight shines on reproducibility, uncomfortable issues will emerge at the interface of research and 'evidence-based' policy.

Consider, for example, the Secret Science Reform Act of 2015, a US bill that would "prohibit the Environmental Protection Agency from proposing, finalizing, or disseminating regulations or assessments based upon science that is not transparent or reproducible". Passed in March by the House of Representatives essentially along party lines (Republicans in favour, Democrats opposed) and now awaiting action by the Senate, the bill has been vigorously opposed by many scientific and environmental organizations.

They argue, probably correctly, that the bill's intent is to block and even roll back environmental regulations by requiring that all data on which the rules are based be made publicly available for independent replication. One of the main objections is that a lot of the scientific research that informs regulatory decisions is not of the sort that can be replicated. For example, a statement of opposition from numerous scientific societies and universities explains that: "With respect to reproducibility of research, some scientific research, especially in areas of public health, involves longitudinal studies that are so large and of great duration that they could not realistically be reproduced. Rather, these studies are replicated utilizing statistical modeling."

Precisely. Replication of the sort that can be done with tightly controlled laboratory experiments is indeed often impossible when you are studying the behaviour of dynamic, complex systems, for example at the intersection of human health, the natural environment and technological risks. But it is hard to see how this amounts to an argument against mandating open access to the data from these studies. Growing concerns about the quality of published scientific results have often singled out bad statistical practices and modelling assumptions, and have typically focused on the very types of science that often underlie regulations, such as efforts to quantify the population-wide health effects of a single chemical.

Although concerns about the bill's consequences are reasonable, the idea that it would be bad to make public the data underlying environmental regulations seems to contradict science's fundamental claims to objectivity and legitimacy. In June, a commentary in *Science* by an array of leading voices, including the current and future heads of the National Academies, flagged "increased transparency" and "increased data disclosure" as crucial elements of science's "self-correcting norm" that can help to address "the disconcerting rise in irreproducible findings" (B. Alberts *et al. Science* **348**, 1420–1422; 2015). This is more or less the position taken by the Secret Science bill's sponsor, Representative Lamar Smith (Republican, Texas): "The bill requires the EPA to use data that is available to the public when the Agency writes its regulations. This allows independent researchers to evaluate the studies that the EPA uses to justify its regulations. This is the scientific method."

This battle for the soul of science is almost surreal in its avoidance of the true issue, which is ideological. One side believes that the government should introduce stricter environmental regulations; the other wants fewer restrictions on the marketplace. Science is the battleground, but it cannot adjudicate this dispute. At its core, the disagreement is about values, not facts. But just as importantly, the facts themselves are inevitably incomplete, uncertain, contested and, as we have been learning, often unreliable.

Like a divorced couple bitterly fighting over the custody of their child, both sides in the Secret Science debate insist that they have only the interests of science at heart. Republicans

are using a narrow, idealized portrayal of science — that it produces clear and reproducible findings — as a weapon to undercut environmental and public-health regulation of the private sector. But many scientists, environmentalists and Democrats have long used similar portrayals to justify the same regulations, and to bash Republicans as anti-scientific when they did not agree.

More and more, science is tackling questions that are relevant to society and politics. The reliability of such science is often not testable with textbook methods of replication. This means that quality assurance will increasingly become a matter of political interpretation. It also means that the 'self-correcting norm' that has served science well for the past 500 years is no longer enough to protect science's special place in society. Scientists must have the self-awareness to recognize and openly acknowledge the relationship between their political convictions and how they assess scientific evidence. ■

**Daniel Sarewitz** is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC. e-mail: [daniel.sarewitz@asu.edu](mailto:daniel.sarewitz@asu.edu)

**QUALITY ASSURANCE WILL INCREASINGLY BECOME A MATTER OF POLITICAL INTERPRETATION.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/t6e7ht](http://go.nature.com/t6e7ht)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## PHYSICS

### A crack in the standard model?

A signal from the Large Hadron Collider (LHC) could challenge the standard model of particle physics for how matter and forces interact.

The LHCb experiment at CERN, Europe's particle-physics lab near Geneva, has uncovered an unexpected difference in the rate at which short-lived particles called *B* mesons undergo certain decays into muons and taus (heavier cousins of the electron). The standard model says that once the particles' mass differences are taken into account, the decays should occur at exactly the same rate.

The deviation is small, and the chance that it is a statistical fluctuation in random noise is too high to claim a discovery (the significance is 2.1 sigma, but physicists' threshold for a discovery is 5 sigma). However, the results are intriguing because they match previous measurements made by two other experiments elsewhere.

*Phys. Rev. Lett.* (in the press)

## IMMUNOLOGY

### Odd fish use old immune trick

Mice and lamprey fish produce a similar antibody response to influenza, despite being separated by hundreds of millions of years of evolution.

Lampreys (pictured) are jawless fish whose common

ancestor with mammals lived 550 million years ago. They defend themselves with antibodies that are unlike those produced by the immune systems of jawed vertebrates. A team led by Jonathan Yewdell at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, exposed lamprey larvae to inactivated influenza virus and found that their blood cells produced antibodies that recognize key amino-acid sites on the head of the haemagglutinin protein of influenza. This is the same region as that targeted by influenza antibodies from mice, suggesting that

lamprey and mouse antibodies recognize pathogens in a similar way despite their huge evolutionary separation. *eLife* 4, e07467 (2015)

## EVOLUTION

### Basque ancestors were farmers

The ancestors of people from the Basque region of Spain were early farmers — not hunter-gatherers as was thought.

Farming practices emerged around 11,000 years ago in the Near East and later spread to Europe as people migrated in waves, eventually replacing the hunter-gatherer lifestyle. To study this influx in Iberia, a

team led by Mattias Jakobsson at Uppsala University in Sweden sequenced the genomes of 8 individuals from remains found in a cave in northern Spain. These people lived 5,500–3,500 years ago, after the arrival of Spain's first farmers around 7,000 years ago. The closest living descendants of the sequenced people are modern Basques, contradicting past studies that linked Basques to late hunter-gatherer groups.

The Basque language is distinct from all other European tongues, and the authors say that it could be a relic of Spain's first farmers. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1509851112> (2015)



## ANIMAL BEHAVIOUR

### Seabirds duped by plastic waste

It is likely that most seabirds have consumed plastic rubbish floating in the ocean after mistaking it for prey.

Chris Wilcox at the Commonwealth Scientific and Industrial Research Organisation in Hobart, Australia, and his colleagues collated published data on the diets of 135 seabird species over the past four decades, including the red-footed booby (*Sula sula*; pictured) and the Cape petrel (*Daption capense*).

According to the data, the proportion of birds that had eaten plastic increased by about 1.7% per year. Using this figure, the team predicts that, had these studies been done today, more than 90% of the seabirds would have eaten plastic. By 2050, that could reach 99% if the flow of plastic waste to the seas is not reduced. The researchers found that the area of highest risk was in the Tasman Sea between Australia and New Zealand. *Proc. Natl Acad. Sci. USA* <http://doi.org/7dv> (2015)

BRITTA DENISE HARDESTY

PAUL VAN HOOF/NATUREINSTOCK.CO/ARDEA.COM



## ECOLOGY

## Predator biomass no match for prey

Twice as much prey does not lead to twice as many predators, according to Ian Hatton at McGill University in Montreal, Canada, and his colleagues.

In theory, more prey should mean more predators. The authors analysed published data on biomass and numbers of individuals for 2,260 ecosystems in 1,512 locations worldwide. They found that the biomass of predators increased along with the biomass of their prey, but that the ratio of the two values decreased. Across ecosystems from grassland to oceans, the ratio scales with an exponent of around 0.75, rather than scaling by 1 as in a linear relationship.

Similar scaling laws are well known between an organism's body mass and features such as metabolism, growth and reproduction, but had not been identified across whole ecosystems, say the researchers. This indicates an unappreciated degree of ecosystem organization.

*Science* <http://doi.org/7f3> (2015)

## ANIMAL PHYSIOLOGY

## How colonies of sea animals swim

The animals that make up a colony of sea creatures contribute to its motion depending on their size.

Complex colonies called physonect siphonophores — relatives of jellyfish — are formed of many individuals that propel the colony using jets of water. John Costello at Providence College in Rhode Island and his team collected colonies of one physonect species (*Nanomia bijuga*) and photographed them as they swam. Individuals worked together to drive the colony around, but did not all contribute equally. Smaller, weaker colony members steered the swimming colony, and their more powerful neighbours

provided the thrust. *Nature Commun.* <http://dx.doi.org/10.1038/ncomms9158> (2015)

## CANCER

## Devil tumour type affects survival

Some lineages of the infectious facial tumours that are devastating populations of the Tasmanian devil (*Sarcophilus harrisii*) can result in worse outcomes for animals.

Rodrigo Hamede at the University of Tasmania in Hobart, Australia, and his colleagues have monitored the outbreak of devil facial tumour disease at a site in northwestern Tasmania since 2006. Animals at this site initially had higher survival rates than other infected populations and a lower proportion of infected animals overall. Their tumours were found to have four sets of chromosomes.

Around 2011–12, this 'tetraploid' tumour lineage was replaced by a 'diploid' type with two sets of chromosomes, which the authors found was associated with an increased disease prevalence in adults (from around 25% of animals infected to 80%) and a significant population decline. Tumour variance can shape both epidemic patterns and outcomes, the authors warn. *Proc. R. Soc. B* 282, 20151468 (2015)

## PHYSICS

## Ring-shaped trap holds ions in check

An electromagnetic trap can suspend 400 ions at a time, providing a useful system for studies of quantum information processing.

Ions can be pinned in place using oscillating and static electric fields, but trapping large numbers is challenging because experimenters must compensate for unavoidable background fields at each ion location. Daniel Stick at Sandia National Laboratories in Albuquerque, New Mexico,

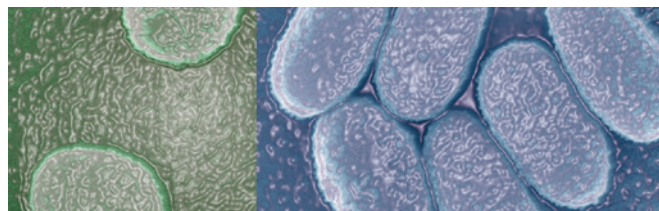
## SOCIAL SELECTION

Popular topics on social media

### Journal of ideas, data and more

With so many journals already in existence, it is rare for a new title to draw attention. But researchers and publishing experts are taking notice of *Research Ideas and Outcomes*, or RIO, an open-access journal that launched on 1 September (<http://rio.pensoft.net>). As well as standard articles, it will publish proposals, experimental designs, data and software, and aims to cover research from all stages of the research cycle. Kelly Visnak, a scholarly-communications librarian at the University of Texas at Arlington, tweeted: "This Open Journal is a game changer." Stephen Curry, a structural biologist at Imperial College London, pondered on Twitter whether open access was now driving "the most innovative & important" developments in publishing.

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/nd2ggr](http://go.nature.com/nd2ggr)



and his colleagues used 88 electrodes to measure and then suppress undesired electric fields at points around a ring-shaped trap. By compensating for background fields in just one direction, they succeeded in trapping 400 calcium ions at uniform intervals around the circle.

Most proposed quantum devices arrange ions in square lattices, but circular chains of ions could find uses in new designs as well as in quantum simulators, say the authors. *Phys. Rev. Appl.* 4, 031001 (2015)

## VIROLOGY

## Giant virus from permafrost

A new species of 'giant' virus has been revived from a 30,000-year-old sample of Siberian permafrost.

The first giant virus visible with light microscopy was seen in 2003. Several species have been discovered since, including *Pithovirus sibericum* (pictured right) found in permafrost in 2014.

Jean-Michel Claverie and Chantal Abergel at the CNRS Institute of Microbiology of the Mediterranean in Marseille, France, and their team have now isolated another giant virus from the same piece of permafrost.

They found the virus — named *Mollivirus sibericum* (pictured left) and seen as spherical particles around 500–600 nanometres in size — multiplying in cultures of amoebas inoculated with the permafrost. Its genome is a double-stranded DNA of 651,523 base pairs, which is unusually devoid of repeats. The authors say that such viruses are probably not rare, and that forms that could infect humans may be reawakened as mining and drilling become more common in the Arctic. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1510795112> (2015)

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

# SEVEN DAYS

The news in brief

## BUSINESS

### Stem-cell safety

Asterias Biotherapeutics reported cautious good news from early-stage trials of an embryonic-stem-cell treatment for the most severe forms of spinal-cord injury on 31 August. Three people with injuries that left them with no feeling from the neck down were injected with low doses of oligodendrocytes — cells derived from embryonic stem cells and that support nerve growth. None experienced serious side effects, and the first participant, operated on in June, showed minor improvements in sensory function. The company, based in Menlo Park, California, plans to increase the number of cells in a dose, hoping to increase the effect.

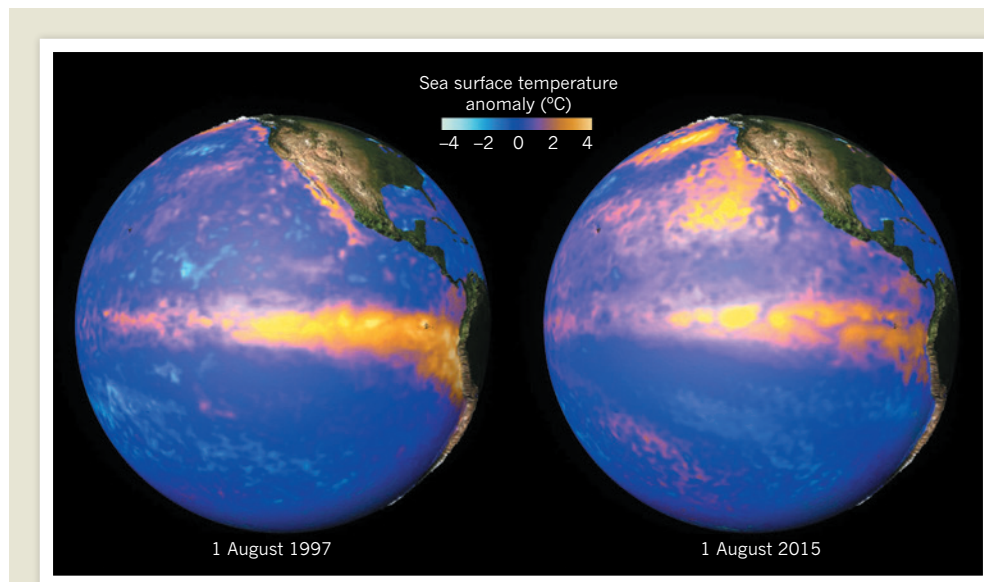
## POLICY

### Ethics revisited

US agencies are working to update ethics rules that regulate biomedical research on humans. On 2 September, the US Department of Health and Human Services (HHS) announced a proposal to revise the 'Common Rule' policy, which governs human-subjects research at 18 US federal departments and agencies. The proposal calls for a single ethical review of research conducted at multiple sites, and more-stringent consent procedures for the use of specimens donated to biobanks. The HHS proposal opened for a 90-day public comment period on 8 September. See [go.nature.com/cbd53s](http://go.nature.com/cbd53s) for more.

### Anthrax review

The US Secretary of the Army has ordered a safety review of all nine Department of Defense laboratories that handle dangerous biological agents.



## El Niño on track to be record-breaking

A comparison of the El Niño weather pattern in 1997 and 2015 shows how the two had developed in a strikingly similar fashion by August in each of the years. Sea-surface temperature data from the National Center for Atmospheric Research in Boulder, Colorado, shows each El Niño as a band of warmer-than-usual water (orange) along the

eastern equatorial Pacific Ocean — with added warmth to the north in 2015. The 1997 event was the strongest in recent memory; the 2015 one now seems as if it could break that record (see *Nature* <http://doi.org/7h6>; 2015). The peak of the event is forecast for late autumn or early winter in the Northern Hemisphere.

The 2 September order follows a problem discovered during an ongoing investigation into a military lab at Dugway Proving Ground in Utah, which in May accidentally shipped live anthrax to a commercial lab. The current investigation found the bacteria outside the normal containment area at Dugway, although still within the confines of the enclosed lab used for dangerous agents. Labs must deliver their safety reports within 10 days of the order.

### Climate appeal

The Dutch government will appeal a landmark ruling on its climate policies, it announced on 1 September. In June, a district court in The Hague declared that the Netherlands must reduce its domestic greenhouse-gas

emissions by at least 25% by 2020, relative to 1990 levels. At present, the Netherlands is on track to achieve a 17% reduction in emissions by 2020, relative to 1990, in line with European Union obligations. A letter from the Ministry of Infrastructure and Environment to the chairman of the House of Representatives argued that the court's verdict might be incompatible with international law. The deadline for appeal is 24 September.

### CRISPR endorsed

Five leading UK research organizations have backed work on human-genome editing. The consortium, which includes the Wellcome Trust and the Medical Research Council, wants to see further debate on the ethics of using

gene-editing technologies such as CRISPR/Cas9, it said in a 2 September statement. These technologies are not yet ready for clinical trials, but the group says that it will continue to fund and support them. In the United Kingdom, genome-editing research is limited at present to non-reproductive cells and human embryos less than 14 days old.

## RESEARCH

### Super Stonehenge

Researchers have discovered a 5,000-year-old row of at least 90 stones at the prehistoric monument of Stonehenge near Salisbury, UK. Found within 3 kilometres of the famous stone circle using non-invasive technologies, some of the stones are as long

VISUALIZATION BY MATT REHME, NATIONAL CENTER FOR ATMOSPHERIC RESEARCH/UCAR



JOHN MOORE/GETTY

as 4.5 metres, the Stonehenge Hidden Landscapes Project announced on 7 September. Its construction could date to about 3,000 BC, the same as Stonehenge. The stones seem to have been purposely buried under the earthworks of the existing Durrington Walls mega-henge. The remnants have not yet been excavated but the team hopes that they will improve understanding of the Neolithic period.

## Forest loss halved

The world's forest area declined at a rate of 3.3 million hectares per year between 2010 and 2015, according to the latest global assessment from the Food and Agriculture Organization (FAO) of the United Nations. However, this is well below the 7.3 million hectares lost per year in the 1990s (R. J. Keenan *et al.* *Forest Ecol. Mgmt* **352**, 9–20; 2015). FAO director-general José Graziano da Silva called the slow-down in deforestation an “encouraging tendency” when he launched the report in Durban, South Africa, on 7 September, but stressed that more still needs to be done.

## AWARDS

## Lasker awards

The 70th Albert and Mary Lasker Foundation awards were announced on 8 September. Evelyn Witkin of Rutgers



University in New Brunswick, New Jersey, and Stephen Elledge of Brigham and Women's Hospital in Boston, Massachusetts, shared the award in basic medical research for their studies into how cells respond to and correct DNA damage. James Allison of the University of Texas MD Anderson Cancer Center in Dallas won the clinical medical research award for work on cancer immunotherapies. The public-service award went to the humanitarian organization Médecins Sans Frontières for its work in battling the 2014 Ebola epidemic in Africa (pictured).

## EVENTS

## Iran deal secured

US President Barack Obama has secured enough Democratic-party votes from the US Senate for July's multilateral deal on Iran's nuclear programme to survive Republican opposition. Republicans are opposed to the deal, and are planning a resolution of disapproval. To stop the resolution completely, the Obama administration needs 41 out of 100 votes in the Senate. But if he gets at least 34 votes in the deal's favour, Obama can veto the disapproval resolution. As of 7 September, 38 Democratic senators supported the deal.

## Polio comeback

Two children have become paralysed in Europe's first polio cases in five years, the World Health Organization reported on 1 September. The cases — in a 10-month-old and a 4-year-old near Ukraine's southwestern borders with Romania, Slovakia, Hungary and Poland — were caused by viruses that are mutated relatives of those in the live polio vaccine. Such vaccine-derived strains are the result of low immunization coverage, but are considered easier to control than outbreaks of wild polio virus. Last year, just half

## COMING UP

### 14–16 SEPTEMBER

Guidelines for using satellite observations to reduce disaster risk are under discussion at the United Nations International Conference on Space-based Technologies for Disaster Management in Beijing.

[go.nature.com/nw3k5r](http://go.nature.com/nw3k5r)

### 14–25 SEPTEMBER

The theory of a world with many dimensions will be debated at length at the Stringy Geometry meeting at the Mainz Institute for Theoretical Physics, Germany.

[go.nature.com/ahdgrg](http://go.nature.com/ahdgrg)

### 15–19 SEPTEMBER

Experts in cell death gather at Cold Spring Harbor, New York.

[go.nature.com/lhszuw](http://go.nature.com/lhszuw)

of Ukrainian children received full immunizations against polio and other preventable diseases.

## Radar retired

One of the two science instruments aboard NASA's US\$916-million Soil Moisture Active Passive (SMAP) satellite was declared dead on 2 September. SMAP was launched in January to produce frequent global maps of soil moisture. But its radar instrument, which measures energy reflecting off Earth's surface, stopped transmitting on 7 July. The problem seems to be with the radar's power-boosting amplifier. SMAP's other science instrument, a radiometer, still works, but losing the radar means that the soil-moisture maps (including freezing and thawing cycles) will be of coarser resolution than planned.

► [NATURE.COM](http://NATURE.COM)

For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

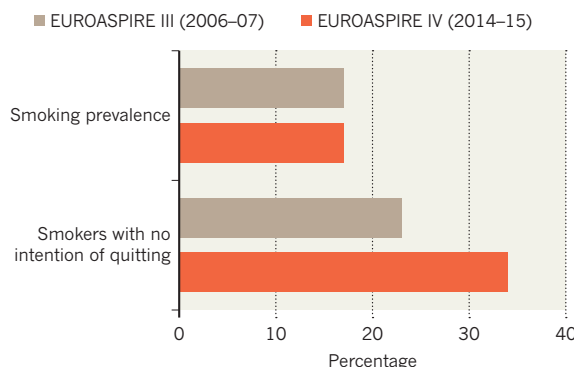
SOURCE: ESC

## TREND WATCH

Two Europe-wide surveys done 8 years apart of people at risk of heart disease show that the proportion who smoke remains the same at 17%. But of those, the proportion who do not intend to stop smoking has risen sharply from 23% to 34%. The EUROASPIRE surveys were run by the European Society of Cardiology in 2006–07 and 2014–15. In total, 5,890 people in Bulgaria, Croatia, Poland, Romania and the United Kingdom were surveyed, and 3,827 participated in both surveys.

## NO INTENTION TO QUIT

Since the last survey in 2006–07, the number of people at highest risk of heart disease who smoke has stayed the same, but the number of those who do not intend to stop has increased.

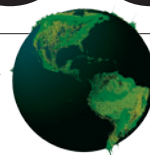


# NEWS IN FOCUS

**CYBERSECURITY** Cryptographers steel themselves for the quantum era **p.167**

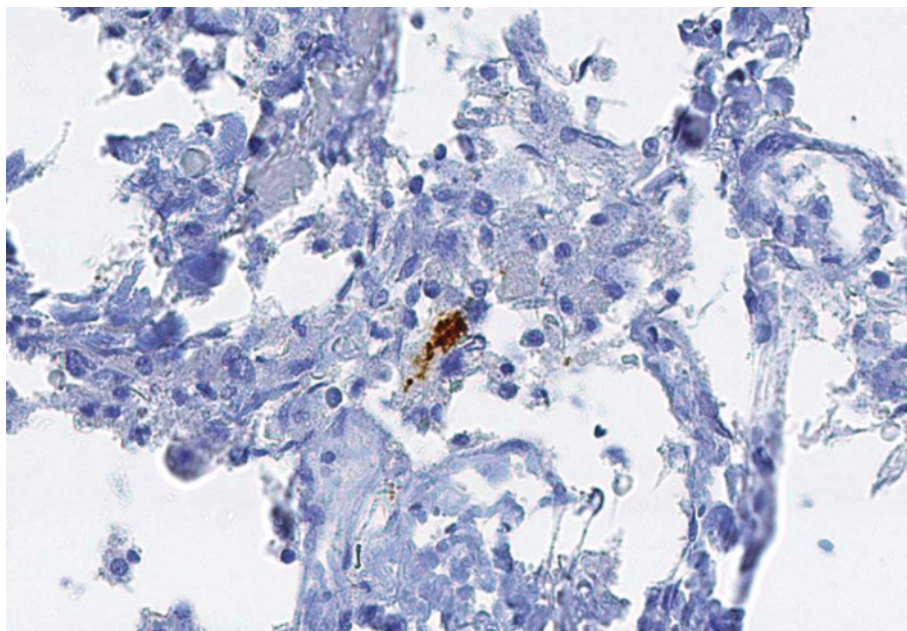
**INSTITUTIONS** German drive for elite universities issues progress report **p.168**

**ECOLOGY** Trees on Earth outnumber stars in the Milky Way **p.170**



**STRUCTURAL BIOLOGY** Cryo-EM reveals a raft of dazzling protein structures **p.172**

JAUNMUKTANE ET AL./NATURE



Amyloid- $\beta$  protein (brown) has been found in the pituitary gland, which sits just outside the brain.

## NEUROSCIENCE

# Alzheimer's fear in hormone patients

*Brain plaques may have been seeded by growth therapy.*

BY ALISON ABBOTT

Only a decade ago, the idea that Alzheimer's disease might be transmissible between people would have been laughed off the stage. But scientists have since shown that tissues can transmit symptoms of the disease between animals — and new results imply that humans, at least in one unusual circumstance, may not be an exception.

The findings, published in this issue of *Nature*, emerged during autopsy studies of the brains of eight people who had died of the rare but deadly Creutzfeldt-Jakob disease (CJD; Z. Jaunmuktane *et al.* *Nature* **525**, 247–250; 2015). They contracted it decades after treatment with contaminated batches of growth

hormone that had been extracted from the pituitary glands of human cadavers. Six of the brains, in addition to the damage caused by CJD, harboured the tell-tale amyloid pathology that is associated with Alzheimer's disease.

"This is the first evidence of real-world transmission of amyloid pathology," says molecular neuroscientist John Hardy of University College London (UCL). "It is potentially concerning."

If confirmed, the findings raise the spectre that tens of thousands of other people treated with the human growth-hormone (hGH) extracts might be at risk of Alzheimer's. And although there is no suggestion that Alzheimer's could be contracted through normal contact with patients, some scientists

worry that the findings may have broader implications: that Alzheimer's could be passed on by other routes through which CJD can be transmitted, such as blood transfusions or contaminated surgical instruments.

CJD is one of several neurodegenerative diseases caused by an infectious, misfolded protein, or prion, called PrP. Its misfolded shape makes it sticky, so it forms clumps. Scientists now believe that Alzheimer's could also be triggered by a similar misfolding, in this case of the peptide amyloid- $\beta$ , with the disease's plaques growing from small amyloid- $\beta$  'seeds'. Mice and marmosets have developed plaques when their brains were injected with brain extracts containing amyloid- $\beta$ ; in mice, plaques developed even when the extracts were injected into the animals' bellies.

The authors of the latest paper provide the first support for the theory that amyloid-plaque formation could be triggered in this way in humans, although "they fall short of providing the final proof of this", says neuroscientist Mathias Jucker of the University of Tübingen, Germany, who is co-author of an accompanying News & Views article (see page 193). Such proof would require injecting the cadaver-derived hGH into animals under controlled conditions and seeing whether amyloid deposits develop as a consequence.

But it may not be easy to get hold of the original hGH extracts, which were prepared in various locations. Some are known to have been stored in Britain, where court cases about possible liability are ongoing, but scientists do not know whether other stocks have been kept. People who received the hGH injections will also be difficult to trace after so many years. The National Prion Clinic at UCL Hospital, which has a helpline for people who are concerned about the risk of CJD after hGH injections, will advise those who call asking about the new developments.

From 1958 until 1985, when the dangers were first realized, around 30,000 people worldwide had hGH injected into their muscles — mostly children who had not been growing at a normal rate. The preparations comprised pooled material extracted from thousands of cadavers. Some extracts turned out to have been contaminated with CJD prions, leading to 226 deadly infections by 2012, mostly in France (119 cases), Britain (65 cases) and the United States (29 cases). Numbers are still creeping up, because CJD has a long incubation period. ►



► None of the eight patients studied, who were aged between 36 and 51 when they died, had shown clinical symptoms of Alzheimer's disease, which also has a long incubation period. Of the six who already had amyloid- $\beta$  pathology, it was widespread in four.

Because it is rare to see this type of amyloid pathology at such young ages, the scientists suspected that amyloid seeds may have been transferred with the hGH injection, just as the CJD prion had been. They did a series of investigations to rule out other explanations.

They determined that none of the eight individuals carried genes that would predispose them to early-onset Alzheimer's or other neurodegenerative diseases. They looked for, but did not find, significant amyloid pathology in patients of a similar age who had died of CJD or other prion diseases but had never been treated with hGH.

Furthermore, the team checked to see whether amyloid pathology really can spread from the brain to the pituitary gland, located just outside the base of the brain. Confirming a 2013 US study, they found that it can spread in principle. They examined the pituitary glands of 49 people who had died with amyloid plaques in their brains and found that seven contained amyloid deposits.

"We think that the most plausible explanation for the occurrence of the amyloid pathology is that it had been transmitted by particular hGH extracts that happened to be contaminated with amyloid- $\beta$  seeds as well as the CJD prions," says John Collinge, a co-author of the paper and a neurologist at UCL. If this turns out to be the case, amyloid- $\beta$  would have been a much more frequent contaminant in the different hGH batches than PrP was, because Alzheimer's is a very common disease.

Prions are harder to deactivate than bacteria and viruses. They stick tightly to metals, and decontamination requires extreme sterilization conditions, which can harm fragile medical instruments. For these reasons, neurosurgeons do not routinely do this type of decontamination, says one German neurosurgeon, speaking off the record — adding that if it were to be confirmed that Alzheimer's is transmitted in a prion-like way, the impact on public health and surgical practice would be major, and very expensive.

"We have learnt a lot about decontamination from our experience with CJD," says neuropathologist Charles Duyckaerts at the Pitié-Salpêtrière Hospital in Paris. "But this is a wake-up call to the medical community to be particularly vigilant."

With so much at stake, scientists are preparing to try to replicate the results independently. Duyckaerts says that he plans to do so on 20 or 30 subjects who died of CJD in France after receiving the cadaver-derived hGH treatment. ■

## BIOMEDICAL SCIENCE

# US agencies plan ethics overhaul

*Government proposes long-awaited revision to regulations designed to protect human subjects.*

BY HEIDI LEDFORD

After years of uncertainty, the US government has revived an effort to update regulations that govern research involving human subjects. The changes would be the most significant since the rules were introduced in 1991.

On 2 September, the US Department of Health and Human Services (HHS) announced a proposal to address concerns that have emerged since the regulations — known collectively as the Common Rule — took effect. These issues include delays caused by overlapping ethics reviews of studies conducted at multiple sites, and the rise of genomic technologies that can identify the donors of anonymized samples.

The HHS will begin a 90-day public-comment period on the proposal next week and will decide how to proceed once that has ended, says Kathy Hudson, deputy director for science, outreach and policy at the US National Institutes of Health (NIH) in Bethesda, Maryland.

The HHS solicited public comments on a similar proposal in July 2011. As the years ticked by without further word on the fate of the revisions, observers grew concerned. "I was totally worried," says Ezekiel Emanuel, a bioethicist at the University of Pennsylvania in Philadelphia, who helped to launch the effort. "It was stalled."

Hudson attributes the delay in part to the need to achieve consensus between the 18 governmental departments and agencies that follow the Common Rule. Research has changed dramatically since the policy was established. Clinical trials are now frequently conducted at multiple sites, with research protocols often reviewed by ethics committees at each place. As a result, it can take a year or more to gain approval for a large, multi-centre trial.

The proposed revisions would authorize a single ethics review for such studies. The NIH plans to enact a similar provision later this year, notes Hudson, but modifying the Common Rule would extend this to other agencies.

The update also suggests simplifying reviews

of research deemed to be of minimal risk to participants. This would reduce the burden on, among others, social scientists who are conducting surveys or collecting oral histories. Emanuel says that this would better protect participants by allowing overtaxed ethics committees to focus their attention on higher-risk research.

Another major provision would require a person's consent to the storing of samples for unspecified future research. At present, such consent is required only when a subject's name or other identifying information is associated with the material. Stripping those data frees the sample for distribution to researchers without consent.

But the rise of genomic sequencing has called into question whether such samples can ever be truly anonymized. Researchers have been able to trace the identities of some subjects on the basis of their DNA sequences. "The people who are participating in research and providing pieces of themselves should be providing permission as well," says Hudson.

That change could put a damper on some research, notes Barbara Koenig, a medical anthropologist at the University of California, San Francisco. "There's a huge public benefit from the research done with de-identified samples," she says. "Requiring explicit consent is going to throw a wrench in that."

Dora Hughes, a senior policy adviser at the law firm Sidley Austin in Washington DC, says that the stricter requirements could also affect the pharmaceutical and medical-device industries. But she commends the HHS for not applying those requirements retroactively to existing samples — a possibility that the department once considered, she says (Hughes is a former HHS counsel). "That discussion raised the spectre of millions of samples that could not be used for research and would otherwise go to waste," she says.

It is not clear how long the HHS will take to finalize the changes, but Hudson says that it is unlikely to wait another four years. She adds that the revision would play an important part in facilitating the planned US Precision Medicine Initiative, a massive government effort to collect genetic, physiological and other health data from 1 million volunteers. "This is really important," Hudson says. "We can't dilly-dally." ■



Today's most widely used encryption methods will not be strong enough resist quantum computers.

## INFORMATION SECURITY

# Encryption faces quantum foe

*Researchers urge readiness against attacks from future-generation computers.*

BY CHRIS CESARE

It is an inevitability that cryptographers dread: the arrival of powerful quantum computers that can break the security of the Internet. Although these devices are thought to be a decade or more away, researchers are adamant that preparations must begin now.

Computer-security specialists are meeting in Germany this week to discuss quantum-resistant replacements for today's cryptographic systems — the protocols used to scramble and protect private information as it traverses the web and other digital networks. Although today's hackers can, and often do, steal private information by guessing

passwords, impersonating authorized users or installing malicious software on computer networks, existing computers are unable to crack standard forms of encryption used to send sensitive data over the Internet.

But on the day that the first large quantum computer comes online, some widespread and crucial encryption methods will be rendered obsolete. Quantum computers exploit laws that govern subatomic particles, so they could easily defeat existing encryption methods.

"I'm genuinely worried we're not going to be ready in time," says Michele Mosca, co-founder of the Institute for Quantum Computing (IQC) at the University of Waterloo in Canada and chief executive of evolutionQ, a

cybersecurity consulting company.

It will take years for governments and industry to settle on quantum-safe replacements for today's encryption methods. Any proposed replacement — even if it seems impregnable at first — must withstand multitudes of real and theoretical challenges before it is considered reliable enough to protect the transfer of intellectual property, financial data and state secrets.

"To trust a cryptosystem, you need a lot of people to scrutinize it and try to devise attacks on it and see if it has any flaws," says Stephen Jordan, a physicist at the US National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland. "That takes a long time."

This week's workshop, held at the Schloss Dagstuhl–Leibniz Center for Informatics in Wadern, is one of several this year bringing together cryptographers, physicists and mathematicians to evaluate and develop cryptographic tools that are less vulnerable to quantum computers. NIST hosted its own workshop in April, and the IQC will team up with the European Telecommunications Standards Institute for another, in early October in Seoul.

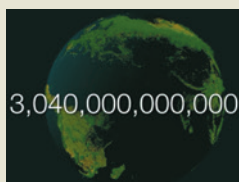
Intelligence agencies have also taken notice. On 11 August, the US National Security Agency (NSA) revealed its intention to transition to quantum-resistant protocols when it released security recommendations to its vendors and clients. And in a memo posted on its website earlier this year, the Dutch General Intelligence and Security Service singled out a looming threat that adds even more urgency to the need for quantum-safe encryption. In a scenario it calls 'intercept now, decrypt later', a nefarious attacker could start intercepting and storing financial transactions, personal e-mails and other sensitive encrypted traffic and then unscramble it all once a quantum computer becomes available. "I wouldn't be at all surprised if people are doing that," says Jordan.

As far back as 1994, mathematician Peter Shor showed that a quantum computer would be able to quickly foil 'RSA encryption', one of the major safeguards used today (P. W. Shor Preprint available at <http://arxiv.org/abs/quant-ph/9508027v2>; 1995). At the time, it was not clear whether such a machine would ever be built, says Mosca, because researchers assumed that it would need to operate flawlessly. But a theoretical discovery in 1996 showed that up to a limit, a quantum computer with some flaws could be just as effective as a perfect one. ►



**MORE  
ONLINE**

## VIDEO HIGHLIGHT



This week's cover, animated  
[go.nature.com/hzuhns](http://go.nature.com/hzuhns)

## MORE NEWS

- LHC signal hints at anomalous particle decay [go.nature.com/29esur](http://go.nature.com/29esur)
- Wikipedia seeks to bridge gap with research community [go.nature.com/oxule4](http://go.nature.com/oxule4)
- Oliver Sacks: an appreciation [go.nature.com/2cloze](http://go.nature.com/2cloze)

## NATURE PODCAST



Thinking differently about autism; plankton in our clouds; and untangling Alzheimer's [nature.com/nature/podcast](http://nature.com/nature/podcast)



► Published experiments with small quantum devices are starting to approach this faultiness threshold, notes Mosca. And because secretive organizations such as the NSA are keenly interested in the technology, it is widely assumed that these published results do not represent the cutting edge of research. “We have to assume there’s going to be people that are a few years ahead of what’s available in the public literature,” says Mosca. “You can’t wait for the headlines in *The New York Times* to have your plan in place.”

The safety of today’s Internet traffic relies in part on a type of encryption called public-key cryptography — which includes RSA — to establish secret communication between users. A sender uses a freely available digital key to lock a message, which can be unlocked only with a secret key held by the recipient. The security of RSA depends on the difficulty of breaking up a large number into its prime factors, which serve as its secret key. In general, the larger the number, the harder this problem is to solve.

Researchers believe that it takes existing computers a long time to factorize big numbers, partly because no one has yet discovered how

to do it quickly. But quantum computers could factorize a large number exponentially faster than any conventional computer, and this nullifies RSA’s reliance on factoring being difficult.

Several options already exist for new public-key cryptosystems. These replace the factoring problem with other difficult mathematics problems that are not expected to yield to quantum computers. Although these systems are not perfectly safe, researchers think that they are secure enough to protect secrets from quantum computers for all practical purposes.

One such system is lattice-based cryptography, in which the public key is a grid-like collection of points in a high-dimensional mathematical space. One way to send a secret message is to hide it some distance from a point in the lattice. Working out how far the encrypted message is to a lattice point is a difficult problem for any computer, conventional or quantum. But the secret key provides a simple way to determine how close the encrypted message is to a lattice point.

A second option, known as McEliece encryption, hides a message by first representing it as the solution to a simple linear algebra problem. The public key transforms the simple problem

into one that seems much more difficult. But only someone who knows how to undo this transformation — that is, who has the private key — can read the secret message.

One drawback of these replacements is that they require up to 1,000 times more memory to store public keys than existing methods, although some lattice-based systems have keys not much bigger than those used by RSA. But both methods encrypt and decrypt data faster than today’s systems, because they rely on simple multiplication and addition, whereas RSA uses more-complex arithmetic.

PQCRYPTO, a European consortium of quantum-cryptography researchers in academia and industry, released a preliminary report on 7 September recommending cryptographic techniques that are resistant to quantum computers (see [go.nature.com/5kellc](http://go.nature.com/5kellc)). It favoured the McEliece system, which has resisted attacks since 1978, for public-key cryptography. Tanja Lange, head of the €3.9-million (US\$4.3-million) project, favours the safest possible choices for early adopters. “Sizes and speed will improve during the project,” she says, “but anybody switching over now will get the best security.” ■

## FUNDING

# Germany claims success for elite universities drive

*Report praises €4.6-billion scheme to make leading universities more competitive — but some smaller institutions have done just as well.*

BY QUIRIN SCHIERMEIER & RICHARD VAN NOORDEN

For a decade, Germany’s government has been trying to explode the myth that all the country’s universities are equal. In 2006, it launched an 11-year, €4.6-billion (US\$5-billion) programme that aimed to make the best German universities more competitive with the likes of Oxford, Cambridge and Harvard. The campaign, called the Excellence Initiative, led to 14 institutions gaining the unofficial label of ‘elite’.

A 3 September report by Germany’s main research-funding agency, the DFG — which administers the initiative together with Germany’s science council — suggests that the cash influx is paying off. Still, a German equivalent of the US Ivy League may be slow to form. An analysis by *Nature*’s news team shows that some universities less favoured by the initiative have improved just as quickly as the elites when

it comes to generating highly cited work. “It doesn’t require the ‘elite’ label to produce good research in Germany,” says Alfred Forchel, president of the University of Würzburg, an institution that has kept pace without top-up funds.

The DFG sees this as positive. “The Excellence Initiative has met expectations,” says Dorothee Dzwonnek, DFG secretary-general. “And it has not weakened universities which don’t directly benefit from it.” But some critics say that the scheme has benefited administrators more than scientists. And a huge increase in research funding across Germany over the past decade makes it difficult to tease out the influence of the initiative on the country’s improvement.

The DFG report, an analysis of funding in German universities that is released every three years, marks the first attempt to measure preliminary outcomes of the initiative. In 2011–13 alone, 45 universities received a total of more than €1 billion for running international graduate schools and setting up specific clusters of

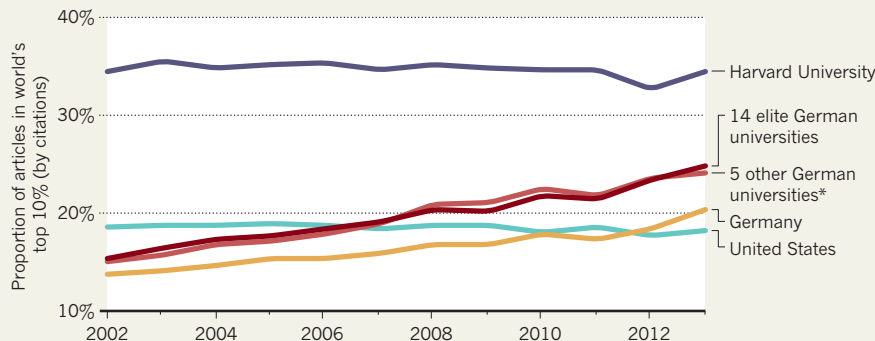
excellence. A subset also each received an extra €10 million to €14 million a year for ‘institutional strategies’ to strengthen the university as a whole — the most prestigious part of the competition (see *Nature* **487**, 519–521; 2012).

The elite group includes some of Germany’s largest and best-equipped research universities, such as the Ludwig Maximilian University of Munich and RWTH Aachen University. The report shows that the elites dominate when it comes to winning competitive grants from the DFG. As a group, they secured more than 40% of the agency’s total funding from 2011–13. However, the same 14 institutions won almost the same share of DFG funding in 2002–04, before the initiative had launched.

Scientific output is booming at the 45 universities that got cash out of the Excellence Initiative, the DFG report points out. They have boosted their output by 43% in chemistry and physics since 2002, more than the 34% increase in these subjects by all German universities.

## GERMANY RISING

Germany's research articles are rapidly becoming more highly cited. But some of the country's smaller universities are matching the rise of its 'elite' institutions.



\*Universities of Bonn, Ulm, Leipzig, Regensburg, Würzburg. Analysis of articles in Scopus database using Elsevier's Scival tool.

found evidence, he says, of booming academic bureaucracy. "For every scientist who has been recruited thanks to the Excellence Initiative, four new administrative positions were created," he says. "It is hardly surprising that elite institutions have no research advantage over some other universities when the group that benefits most from the Excellence Initiative is not involved in science."

The DFG says that it has not looked at how the programme may have burdened university administrations. "It has attracted 4,000 talented foreign scientists to German universities and it has greatly increased these universities' scholarly output," says Dzwonnek. "From our point of view, this is a real success." Many agree that the competition, despite ambiguity over its measurable impacts, has served German science well. It was a positive shock to Germany's structurally conservative science system, says Jakob Edler, executive director of the Manchester Institute of Innovation Research, UK.

The results of a comprehensive evaluation of the Excellence Initiative by an international panel of experts are due in January 2016. The federal government and Germany's 16 states, which have tentatively agreed to continue the programme, will then decide about its future. "The Excellence Initiative promotes fresh ideas and new collaborations. I do hope it continues beyond 2017," Forchel says. ■

And a further analysis by *Nature* finds that the 14 elites alone now produce 35% of Germany's total articles, up from a share of 29% in 2002.

### THE REST CHASE THE BEST

But the Excellence Initiative may not be separating the elites from the rest when it comes to the quality of research papers. *Nature's* analysis shows that almost one-quarter of articles from the elites are now in the world's top 10% by citations — up from one-sixth 12 years ago. Yet it also shows that some other German universities

that received much less funding, or no top-up funds, have matched this rise (see 'Germany rising'). That is enlightening, says Karl Ebeling, president of the University of Ulm, which had little success in the initiative but is higher in some international rankings than elite universities in Bremen and Konstanz, for instance.

Björn Brembs, a neurobiologist at the University of Regensburg, thinks that the unclear impact of the initiative on creating elites is because the cash was poorly spent. He has delved into German employment statistics and



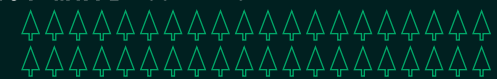
# TRILLIONS OF TREES

SURVEY OF SURVEYS FINDS 422 TREES FOR EVERY PERSON ON EARTH

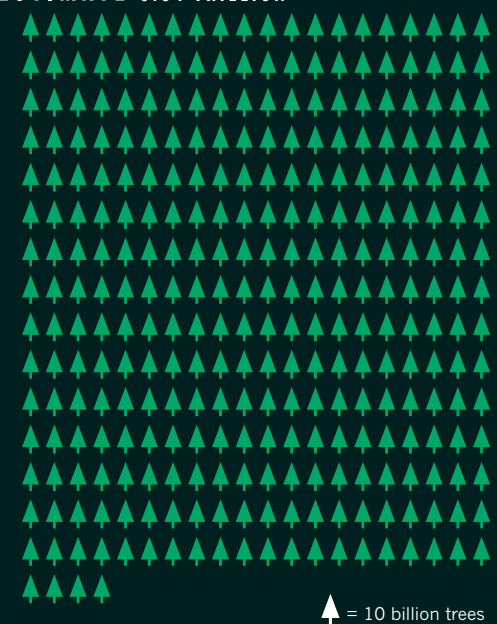
BY RACHEL EHRENBURG  
DATA VISUALIZATION BY JAN WILLEM TULP

Three trillion: the latest estimate of the planet's tree population, published in this issue of *Nature* (see page 201), exceeds the number of stars in the Milky Way. At more than 7 times the previous estimate of 400 billion, the figure is impressive, but it should not necessarily be taken as good news. The forest-density study — which combined satellite imagery with data from tree counts on the ground that covered more than 4,000 square kilometres — also estimated that 15 billion trees are cut down each year. And in the 12,000 years since farming began spreading across the globe, the number of trees on our planet has fallen by almost half.

## OLD ESTIMATE 400 BILLION

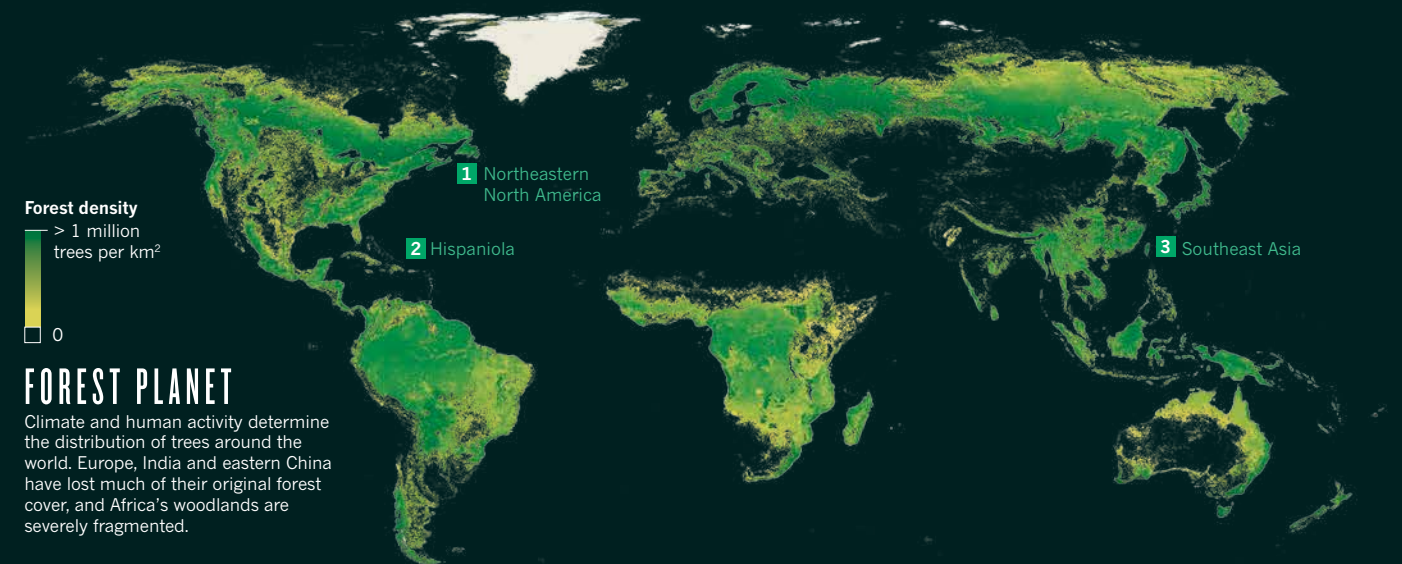


## NEW ESTIMATE 3.04 TRILLION



**NATURE.COM**  
For an animated data visualization, visit:  
[go.nature.com/h8ucmu](http://go.nature.com/h8ucmu)

Line height represents forest density in 1 km<sup>2</sup>



## FOREST PLANET

Climate and human activity determine the distribution of trees around the world. Europe, India and eastern China have lost much of their original forest cover, and Africa's woodlands are severely fragmented.

### 1. NORTHEASTERN NORTH AMERICA

Farms, orchards and sheep took over the landscape of northeastern North America in the 1800s, when much of the region's forest was harvested for timber. Today, the six US states of New England are more than 80% forested — but suburban sprawl and other factors present new threats.



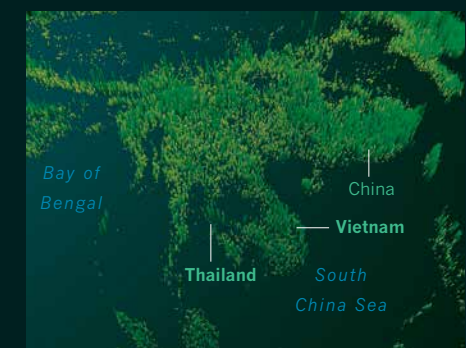
### 2. HISPANIOLA

The effects of deforestation are stark on the Caribbean island of Hispaniola. The Dominican Republic, on the eastern side of the island, has tree cover that is four times denser than that in neighbouring Haiti, which has been forced to cut down trees for fuel.



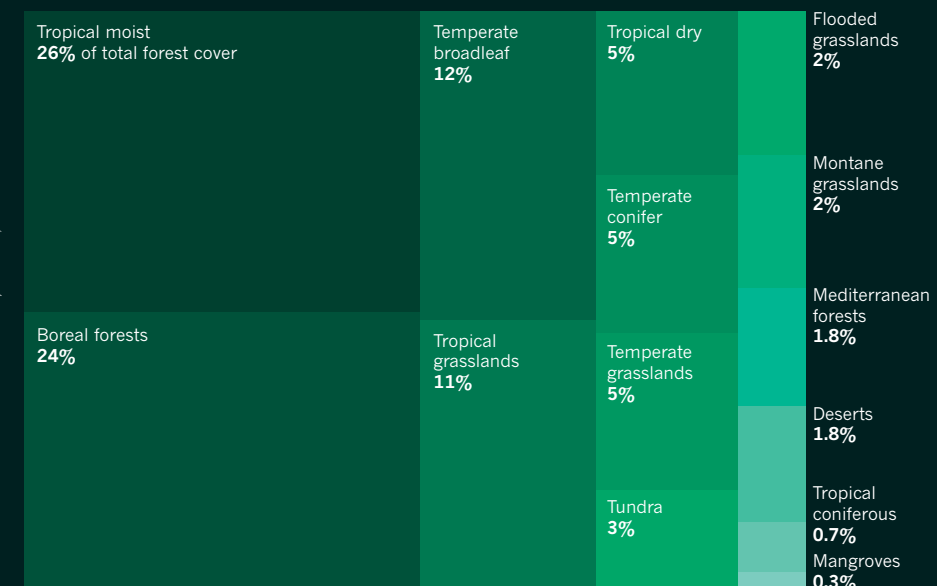
### 3. SOUTHEAST ASIA

Forests in southeast Asia have changed drastically since the 1970s. From 1973 to 2009, Thailand and Vietnam lost 43% of their forest cover; Cambodia and Laos lost 22% and 24%, respectively. If current trends continue, more than 30% of the region's remaining forest will be cleared by 2030.



## LAY OF THE LAND

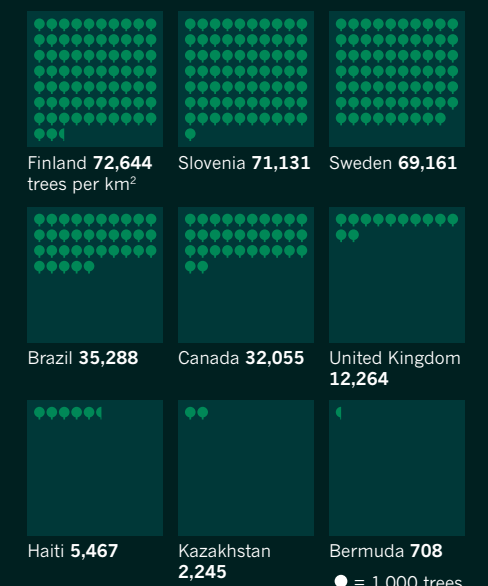
Despite deforestation caused by farming, ranching, mining and logging, tropical areas still contain an astounding 43% of the planet's trees. Tree densities are greatest in the northern boreal and tundra forests, which can contain more than 1,000 trees per hectare. (Percentages are rounded.)



SOURCE: T. W. CROWTHER ET AL. NATURE 525, 201–205, 2015

## LEAF OF NATIONS

The tropics host many densely forested countries, but nations with boreal forest, such as Finland, have the highest tree densities. At the other extreme are desert and island nations, and some impoverished countries.



● = 1,000 trees





# THE REVOLUTION WILL NOT BE CRYSTALLIZED

**MOVE OVER X-RAY CRYSTALLOGRAPHY.  
CRYO-ELECTRON MICROSCOPY IS  
KICKING UP A STORM IN STRUCTURAL  
BIOLOGY BY REVEALING THE HIDDEN  
MACHINERY OF THE CELL.**

**BY EWEN CALLAWAY**

**I**n a basement room, deep in the bowels of a steel-clad building in Cambridge, a major insurgency is under way.

A hulking metal box, some three metres tall, is quietly beaming terabytes' worth of data through thick orange cables that disappear off through the ceiling. It is one of the world's most advanced cryo-electron microscopes: a device that uses electron beams to photograph frozen biological molecules and lay bare their molecular shapes. The microscope is so sensitive that a shout can ruin an experiment, says Sjors Scheres, a structural biologist at the UK Medical Research Council Laboratory of Molecular Biology (LMB), as he stands dwarfed beside the £5-million (US\$7.7-million) piece of equipment. "The UK needs many more of these, because there's going to be a boom," he predicts.

In labs around the world, cryo-electron microscopes such as this one are sending tremors through the field of structural biology. In the past three years, they have revealed exquisite details of protein-making ribosomes, quivering membrane proteins and other key cell molecules,

ILLUSTRATION BY VIKTOR KOEN



discoveries that leading journals are publishing at a rapid clip. Structural biologists say — without hyperbole — that their field is in the midst of a revolution: cryo-electron microscopy (cryo-EM) can quickly create high-resolution models of molecules that have resisted X-ray crystallography and other approaches, and labs that won Nobel prizes on the back of earlier techniques are racing to learn this upstart method. The new models reveal precisely how the essential machinery of the cell operates and how molecules involved in disease might be targeted with drugs.

“There’s a huge range of very important biological problems that are now open to being tackled in a way that they could never before,” says David Agard, a structural cell biologist at the University of California, San Francisco.

Scheres was recruited to the LMB several years ago to help push cryo-EM technology to its limits — and he and his colleagues have done just that. Last month, they reported one of the burgeoning field’s most impressive feats: a startlingly clear picture of an enzyme implicated in Alzheimer’s disease, showing the position of its 1,200 or so amino acids down to a resolution of a few tenths of a nanometre<sup>1</sup>.

Biologists are now pushing the technique further to deduce ever more detailed structures of small and shape-shifting molecules — a challenge even for cryo-EM. “Whether you call it revolution or a quantum leap, the fact is that the gates have opened,” says Eva Nogales, a structural biologist at the University of California, Berkeley.

## CRYSTAL COAXING

Spend a bit of time with a structural biologist and they will probably mention their field’s unofficial motto: ‘structure is function’. Only by knowing the atom-by-atom arrangement of a biomolecule can researchers grasp how it works — how, for instance, the ribosome reads strands of messenger RNA to manufacture proteins, or how molecular pores flip open and shut. For decades, one technique enjoyed a near monopoly in elucidating protein structures to this level of detail: X-ray crystallography, in which scientists persuade proteins to form into crystals, then blast X-rays at them and decipher the protein’s structure from patterns that the X-rays make when they bounce off (see ‘Structure solvers’). Of the more than 100,000 entries in the Protein Data Bank, a popular repository of protein structures, about 90% were solved by this technique. It has contributed to more than a dozen Nobel prizes, including the one awarded in 1962 for revealing DNA’s double helix.

But although X-ray crystallography has been structural biologists’ best tool, it also has major limitations. It can take researchers years to find ways of forming some recalcitrant proteins into large crystals that are suitable for analysis, and many fundamentally important molecules — such as proteins that are embedded in cell membranes or that make up complex molecular machines — have defied crystallization.

X-ray crystallography was certainly king when biologist Richard Henderson arrived at the LMB in 1973 to study a protein called bacteriorhodopsin, which uses light energy to pump protons across a membrane. Henderson and his colleague Nigel Unwin had managed to make two-dimensional crystals from the protein, but they were unsuitable for X-ray diffraction. So the pair decided to try electron microscopy instead.

At the time, electron microscopy was used to study viruses or slices of tissue that had been treated with heavy-metal stains. A beam of electrons is fired at a sample, and the emerging electrons are detected and used to map out the structure of the materials they smashed into. This approach produced the first detailed image of a virus — a tobacco pathogen — but the stain made it difficult to see individual proteins, let alone the atomic details that the X-rays were revealing. “It was blobby stuff or negative-stained, and you would see outlines of molecules,” says Agard.

In a pivotal step, Henderson and Unwin omitted the stain when they used electron microscopy to image crystal sheets of bacteriorhodopsin — instead, they placed the crystals on metallic grids to make the protein stand out. “You were looking at the atoms in the protein,” says Henderson, who, with Unwin, published<sup>2</sup> the structure of bacteriorhodopsin in 1975. “That was

such a huge step forward,” Agard says. “That said, ‘OK, it will be possible to solve protein structures by EM.’”

The cryo-EM field developed through the 1980s and 1990s; a key advance was the use of liquid ethane to flash-freeze proteins in solution and hold them still<sup>3</sup>, which is how the ‘cryo’ came to cryo-EM. But still the technique could generally resolve structures only to more than 10 Ångströms (1 Å is one-tenth of a nanometre) — nothing to rival the better than 4-Å models of X-ray crystallography, and nowhere near what was needed to use the structures for drug design. While funders such as the US National Institutes of Health were ploughing hundreds of millions of dollars into ambitious crystallography initiatives, support for cryo-EM lagged far behind.

In 1997, when Henderson attended the annual Gordon Research Conference on 3D electron microscopy, a colleague opened the meeting with a provocative statement: cryo-EM was a “niche” method, he said, unlikely to ever supplant X-ray crystallography. But Henderson could see a different future, and he fired back a salvo in the next talk. “I said we should go for global domination of cryo-EM over all the structural methods,” he recalls.

## THE REVOLUTION STARTS HERE

In the years that followed, Henderson, Agard and other cryo-EM evangelists worked methodically on technical improvements to electron microscopes — in particular, on better ways to sense electrons. Long after digital cameras had taken the world by storm, many electron microscopists still preferred old-fashioned film because it recorded electrons more efficiently than did digital sensors. But, working with microscope manufacturers, the researchers developed a new generation of ‘direct electron detectors’ that vastly outperforms both film and digital-camera detectors.

Available since about 2012, the detectors can capture quick-fire images of an individual molecule at dozens of frames per second. Researchers such as Scheres, meanwhile, have written sophisticated software programs to morph thousands of 2D images into sharp 3D models that, in many cases, match the quality of those deciphered with crystallography.

Cryo-EM is suited to large, stable molecules that can withstand electron bombardment without jiggling around — so molecular machines, often built from dozens of proteins, are good targets. None has proved more suitable than ribosomes, which are braced by rigid twists of RNA. The solution of ribosome structures by X-ray crystallography won three chemists the 2009 Nobel Prize in Chemistry — but those efforts took decades. In the past couple of years, ‘ribosomania’ has gripped cryo-EM researchers, and various teams have quickly determined and published dozens of cryo-EM structures of ribosomes from a multitude of organisms, including the first high-resolution models of human ribosomes<sup>4,5</sup>. X-ray crystallography has largely fallen by the wayside in the LMB laboratory of Venki Ramakrishnan, who shared the 2009 Nobel. For large molecules, “it’s safe to predict that cryo-EM will largely supersede crystallography”, he says.

The rocketing number of cryo-EM publications suggests this to be true: in 2015 alone, the technique has so far been used to map the structures of more than 100 molecules. And, unlike X-ray crystallography, in which crystals lock proteins in a single, static pose, researchers can use cryo-EM to calculate the structure of a protein that has been flash-frozen in several conformations and so deduce the mechanisms by which it works.

In May, structural biologist John Rubinstein at the University of Toronto, Canada, and his colleagues used around 100,000 cryo-EM images to create a ‘molecular movie’ of a rotor-shaped enzyme called V-ATPase, which pumps protons in and out of cell vacuoles by burning ATP<sup>6</sup>. “What we saw is that everything is flexible,” Rubinstein says. “It’s bending and twisting and deforming.” He thinks that the enzyme’s flexibility helps it to efficiently transmit energy released by ATP to the pump.

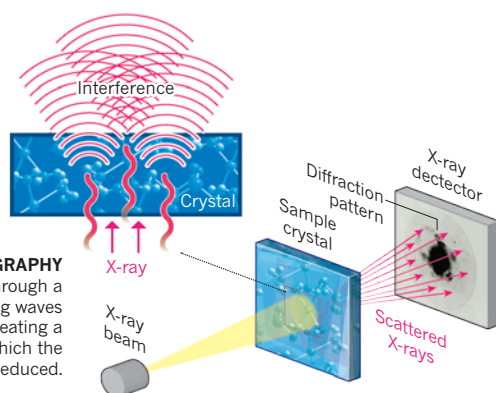
## “WE SHOULD GO FOR GLOBAL DOMINATION OF CRYO-EM OVER ALL THE STRUCTURAL METHODS.”

➔ **NATURE.COM**  
See a selection of  
stunning cryo-EM  
structures at:  
[go.nature.com/cehow8](http://go.nature.com/cehow8)

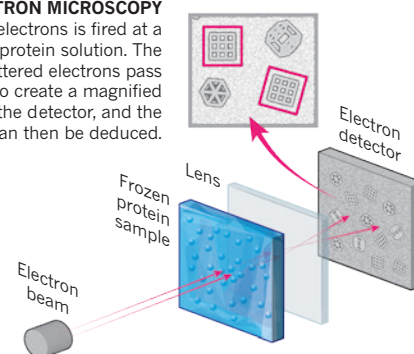
## STRUCTURE SOLVERS

X-ray crystallography has long been the dominant method for deducing high-resolution protein structures, but cryo-electron microscopy is catching up.

**X-RAY CRYSTALLOGRAPHY**  
X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.



**CRYO-ELECTRON MICROSCOPY**  
A beam of electrons is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, and the structure can then be deduced.



X-RAY IMAGE: SPL

And when a team led by Nogales in 2013 pieced together cryo-EM images of a complex that orchestrates the transcription of DNA into RNA, they discovered that an entire arm swings 100 Å around the DNA strand like a crane, potentially influencing whether a gene is transcribed<sup>7</sup>. “I think this is beautiful,” says Nogales. “It’s a true insight into how these biological machines work.”

## SMALL AND BEAUTIFUL

Now that cryo-EM has hit its stride, experts are looking for grander challenges. For many, the most coveted targets are smaller proteins sandwiched in cellular membranes. These tend to be linchpins in cellular signalling pathways, as well as popular drug targets. They are also notoriously difficult to crystallize, and imaging individual proteins with cryo-EM is tough because it is harder to extract the signal from the background noise.

These hurdles did not stop Yifan Cheng, a biophysicist at the University of California, San Francisco (UCSF), from attempting cryo-EM on a small membrane protein called TRPV1, which detects the molecule that gives chilli peppers their burn and is closely related to other pain-sensing proteins. A team led by his collaborator David Julius, a UCSF physiologist, had failed to crystallize the protein. The cryo-EM project was slow-going at first, but the same technical advances that drove ribosomania produced a 3.4-Å structure of TRPV1 in late 2013. The report<sup>8</sup> was a thunderbolt to the field, because it showed that cryo-EM could conquer small, medically important molecules. “I literally lost an entire night’s sleep when I saw that,” says Rubinstein.

More sleepless nights are likely to follow. “There’s going to be a huge explosion in the number of membrane-protein structures that get solved,” says Agard.

One such solution was that published last month<sup>1</sup> by Scheres, structural biologist Yigong Shi of Tsinghua University in Beijing and their team. They produced a model of γ-secretase — a protein that makes the amyloid-β molecule that is linked to Alzheimer’s disease. The 3.4-Å-resolution map reveals that γ-secretase mutations that cause rare inherited forms of Alzheimer’s map to two ‘hotspots’ in the enzyme and seem to influence its ability to form toxic amyloid-β particles. The structure could help researchers to understand why drugs that inhibit the enzyme have failed in past clinical trials, and help them to design new pharmaceuticals. “Stunning” is how Cheng describes the structure.

Results such as these are attracting the attention of drug companies hoping to study medically important proteins that have resisted crystallography. Scheres is working with New York-based pharmaceutical giant Pfizer on ion channels, a broad class of membrane protein that includes pain-sensing molecules and neurotransmitter receptors. “I’ve been contacted by almost everybody,” says Nogales of the drug companies lining up at her door.

But despite the advances, many in the field see room for further improvement. They hope to devise better electron detectors and better methods for preparing protein samples. This would allow scientists to image proteins that are even smaller and more dynamic, and at even greater resolution than before. A 2.2-Å structure of a bacterial enzyme

published in May<sup>9</sup> showed just how sharp cryo-EM structures can get.

Like any burgeoning field, this one has growing pains. Some experts worry that researchers rushing to use the technique could produce problematic results. A 2013 structure of an HIV surface protein<sup>10</sup> was questioned by scientists who said that the images used to build the model were white noise<sup>11</sup>. Since then, X-ray and cryo-EM models generated by other teams have challenged the original model, but the researchers have stood by their result<sup>12</sup>. This June, at the field’s Gordon conference, researchers wanting more quality control passed a resolution urging journals to provide referees with details of how cryo-EM structures were created.

Costs could slow the spread of the technology. Scheres estimates that the LMB spends around £3,000 per day running its cryo-EM facility, plus another £1,000 on electricity, most of it for computers needed to store and process the images. “You’re £4,000 per day lighter if you want to do this. That, for many places, is a very high cost,” he says. To make cryo-EM more accessible, some funders have established shared facilities at which researchers can book time. The Howard Hughes Medical Institute (HHMI) operates a cryo-EM lab on its Janelia Farm Campus in Virginia that is open to HHMI-funded investigators based elsewhere. In the United Kingdom, a national cryo-EM facility funded by the government and the Wellcome Trust opened this year in Didcot, near Oxford. “There is a real tidal wave of people wanting to learn about it,” says Helen Saibil, a structural biologist at Birkbeck, University of London, who helped to establish the UK facility.

Riding the wave is Rod MacKinnon, a biophysicist at Rockefeller University in New York City, who shared the 2003 Nobel Prize in Chemistry for determining the crystal structure of certain ion channels, but who is now deep into cryo-EM. “I’m on a steep slope of a learning curve, which always thrills me,” says MacKinnon, who hopes to use the method to study how ion channels open and close.

Henderson’s tongue may have been firmly in his cheek when he declared back in 1997 that cryo-EM could rule the structural-biology world. But nearly 20 years later, his prediction is looking less like hyperbole than it did then. “If it carries on, and all the technical problems are solved, cryo-EM could indeed become, not just a first choice, but a dominant technology,” he says. “We are probably halfway there.” ■

Ewen Callaway writes for Nature from London.

1. Bai, X.-C. *et al. Nature* <http://dx.doi.org/10.1038/nature14892> (2015).
2. Unwin, P. N. T. & Henderson, R. *J. Mol. Biol.* **94**, 425–440 (1975).
3. Adrian, M., Dubochet, J., Lepault, J. & McDowell, A. W. *Nature* **308**, 32–36 (1984).
4. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. *Science* **348**, 95–98 (2015).
5. Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. *Nature* **520**, 640–645 (2015).
6. Zhao, J., Benlekhir, S. & Rubinstein, J. L. *Nature* **521**, 241–245 (2015).
7. Cianfrocco, M. A. *et al. Cell* **152**, 120–131 (2013).
8. Liao, M., Cao, E., Julius, D. & Cheng, Y. *Nature* **504**, 107–112 (2013).
9. Bartesaghi, A. *et al. Science* **348**, 1147–1151 (2015).
10. Mao, Y. *et al. Proc. Natl Acad. Sci. USA* **110**, 12438–12443 (2013).
11. Henderson, R. *Proc. Natl Acad. Sci. USA* **110**, 18037–18041 (2013).
12. Mao, Y., Castillo-Menendez, L. R. & Sodroski, J. G. *Proc. Natl Acad. Sci. USA* **110**, E4178–E4182 (2013).





# Fishing for the first Americans

*Archaeology is moving underwater and along riverbanks to find clues left by the people who colonized the New World.*

BY EMMA MARRIS

**O**n 17 September, a catamaran will set off into the Pacific Ocean on a week-long cruise back to the Pleistocene. Laden with sonar instruments, the research vessel *Shearwater* will probe the ocean bottom to find places that were beaches and dry land more than 13,000 years ago, when the sea level was around 100 metres lower. The researchers are hunting for evidence that ancient people lived along this now-sunken coastline as they colonized the New World.

Meanwhile, other archaeologists are digging in the intertidal zone on a remote island off the shore of British Columbia in Canada, where the sea level has barely changed since the ice-age glaciers began to retreat. Since late last year, that team has found footprints and a tool that date back 13,200 years, making them some of the oldest human marks on



## Evidence of ancient Americans has turned up on Calvert Island in Canada.

the continent. Whoever left them had to have reached the island by boat.

Welcome to the newest wave of American archaeology: the idea that the first residents of the Americas came by sea, hugging the Pacific coast as they went south. This theory marks a sharp departure from the once-dominant hypothesis that Pleistocene hunters from Siberia migrated by foot across a land bridge to Alaska and then south into the heart of North America. This route opened up only when the vast sheets of ice covering the continent had melted enough to permit passage. It was thought that these first migrants made the distinctive stone spear tips called Clovis points, which began appearing at sites in the interior of North America around 13,000 years ago.

There has long been evidence that others reached the New World at least 1,000 years earlier. But only in the past decade have archaeologists accumulated enough evidence to abandon the Clovis-first model (see *Nature* **485**, 30–32; 2012). Some of the earliest human sites in the Americas date to well before a corridor opened up between the ice sheets, which is forcing researchers to explore the idea that New World colonizers skirted the coastline. Travelling by boat, these early people could have hoppedscotched their way south of the ice sheets, subsisting on the rich marine resources of the ice-free strip along the shore.

The search for these sea-going settlers will not be easy. Much of the evidence that archaeologists seek is deep underwater — or was smashed long ago by the Pacific's legendary waves. But momentum is building to find those earliest settlers. "People are just more optimistic," says Quentin Mackie, an archaeologist at the University of Victoria in Canada. Amanda Evans, a marine archaeologist at the ocean-survey company Tesla Offshore in Prairieville, Louisiana, says that prehistoric underwater archaeology in general is having a moment. "This year just seems to be the year that everybody was pushing the ball uphill and it finally crested."

### TOOLS OF THE TRADE

Loren Davis, an archaeologist at Oregon State University in Corvallis, is searching for the ancient seafarers in an unusual spot — at a site in Idaho called Cooper's Ferry, which is on a bank of the Salmon River, hundreds of kilometres away from the coast. At the dig site in August, Davis examines a piece of rock brought to him by one of his field crew. He turns it over to see whether it was shaped by human hands, perhaps by early toolmakers who littered the ground with flakes of rock as they worked.

Although Cooper's Ferry is far inland, Davis suggests that it is part of the coastal story. The Salmon is a tributary of the mighty Columbia River, which would have been the first large



Archaeologists search for signs of early inhabitants near a riverbank at Cooper's Ferry, Idaho.

waterway encountered by people who made it south of the ice sheets during the last glacial epoch. At that time, valleys farther north would have been covered by glaciers. For a water-adapted culture, he says, "the first off-ramp south of the ice is the Columbia River".

Having considered the stone, Davis hands it back to his colleague and says, "I think it is a flake." His archaeological pits, which the crew has shaped into a series of neat holes, are full of flakes and finished 'western stemmed' points up to 13,200 years old<sup>1</sup>. Whereas the Clovis points are shaped like miniature surfboards, the western stemmed points from Cooper's Ferry are smaller and look like Christmas trees. Points resembling the western stemmed variety have been found throughout the western United States and in Siberia — a connection that suggests they were brought over to the New World by early hunters.

first Americans carried these points there by sea and river. "You get a gambler's mentality," Davis says. The hunt obsesses the crew, who spend weeks here, camped out and digging for hours each day. Sarah Skinner, an Oregon State student who supervises pit B, says that she wakes up clenching her fists around dream trowels. "When I close my eyes, I see artefacts," she says.

### HINGE-POINT HUNT

Signs of early inhabitants are also starting to appear along the coast, particularly in spots where the swelling seas have not covered ancient shorelines since the end of the last glacial period. The western coast of Canada, for example, was pressed down by the Pleistocene ice and has been rebounding upwards since the glaciers melted. In some spots — hinge points — that post-glaciation rebound almost exactly cancels out the rising sea level<sup>2</sup>. One of

*"This is probably the biggest effort to identify submerged sites along the Pacific coast."*

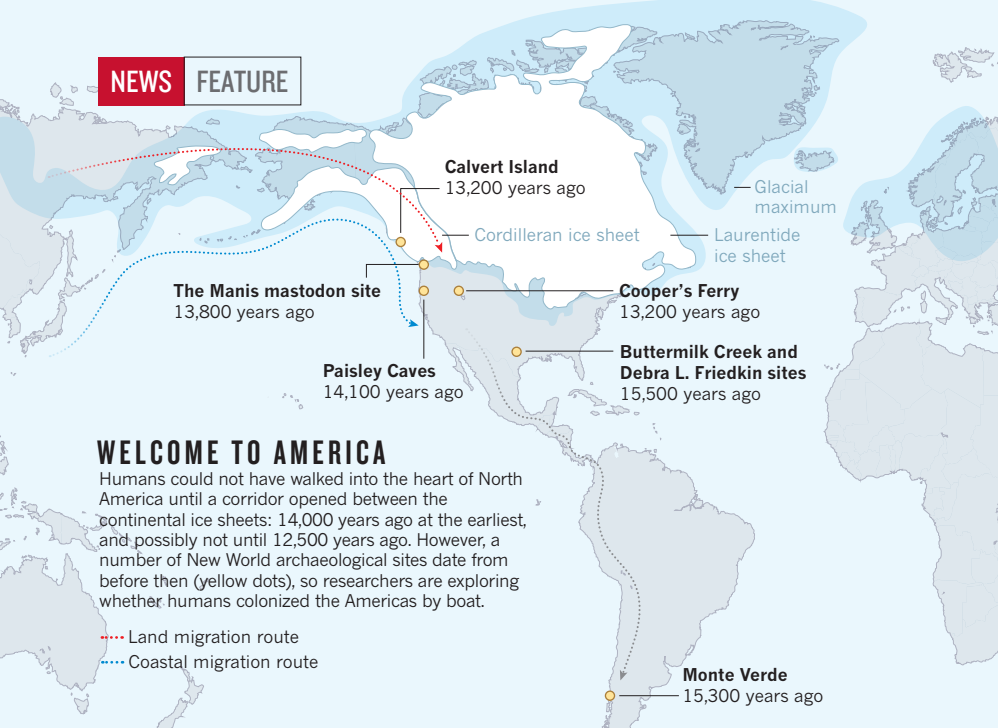
Davis's crew is quietly intent, and the air is filled with the gentle sound of trowels scraping earth, along with a rock wren's distinctive call. The peace is occasionally broken by shouts between diggers and data recorders: "Bone!", "Fire-cracked rock!" or "Debl!" (short for debitage, or flakes). The position of each artefact is precisely recorded, then it is bagged up and stored in one of many boxes that are piling up in a nearby trailer. Precise dates will be assigned later, in the laboratory.

A sense of expectation hangs over the dig. If the team uncovers particularly old western stemmed points that definitively pre-date the Clovis era, that would strongly suggest that the

those hinge points is Calvert Island, where a 13,200-year-old footprint was found late last year and another was discovered this summer. Daryl Fedje and Duncan McLaren, both archaeologists at the University of Victoria in British Columbia, plan to continue working the site to look for signs of the earliest Americans (see 'Welcome to America').

The Hakai Institute on Calvert Island, founded by Canadian entrepreneur Eric Petersen, is supporting that work. "As a fourth-generation British Columbian," Peterson says, "I am intensely interested in the rich history of humans on our coastline, which we now realize goes way, way back. How far back? Thirteen





thousand years? Fifteen thousand years? That's what we want to find out."

Mackie thinks that Calvert Island and similar hinge points will produce results much faster than underwater work, which is technically challenging and expensive. "You might as well just stand on your boat and burn \$100 bills," he says.

But despite the enormous cost and technical challenges, he and others agree that underwater locations may hold tremendous potential and that the time has finally come for archaeologists of American prehistory to explore the Pacific Ocean. There have been a few projects over the years. In the late 1990s, Mackie and Fedje did some sea-floor mapping around Haida Gwaii, an island off the British Columbia coast<sup>3</sup>. They took samples of sea-floor sediment and hauled up a barnacle-encrusted flake tool that they suggest dates from 10,000 years ago, when the sea floor on which it was found was dry land. More recently, the duo used an autonomous underwater vehicle to explore and found what they suggest could be a fishing weir — a trap made from rocks — that dates back 13,800 years.

Archaeologist James Dixon of the University of New Mexico in Albuquerque has done marine surveys of the now-submerged land that once connected Asia and North America. And Jon Erlandson, an archaeologist at the University of Oregon in Eugene, has worked for years on the Channel Islands off Southern California, piecing together evidence for his theory that people followed a 'kelp highway' down the coast. This route would have offered abundant food — fish, shellfish and marine mammals — supported by the kelp forest ecosystems.

But the offshore studies so far have been limited, and most of the discoveries have not been reliably dated. "There's very few of us, and we are spread over vast, vast areas," says Dixon. There has been a hesitation to join — or fund — the chancy and expensive underwater search. "You can go out there and be totally skunked by the

weather," he says. "It is a tough thing."

That has led many researchers to discount the coastal hypothesis in the past, notes Mackie. "People thought 'well, all the information is deep underwater and we'll never find it,'" he says. And so far, nothing older than the Clovis era has been found along the Pacific Coast — either on the sea floor or on land.

### SPARK OF INTEREST

What snapped the field out of its funk was not a charismatic leader or a spectacular find — it was federal bureaucracy.

The US Bureau of Ocean Energy Management (BOEM) was formed in 2010 to regulate energy development on the continental shelf. The bureau is bound by the National Historic Preservation Act, which requires it to make sure that valuable archaeological sites will not be destroyed by any development that requires a federal permit. As interest in offshore renewable-energy projects has increased in recent years, BOEM has scrambled to improve methods for identifying prehistoric sites.

In 2011, it commissioned a sweeping study of possible archaeological sites on the Pacific outer continental shelf. Davis and a colleague at Oregon State, archaeologist Alex Nyers, worked on the report, using existing ocean-depth data and estimates of sea-level rise to decipher where previous shorelines would have been<sup>4</sup>. They then modelled where prehistoric sites might be clustered: presumably on gentle, south-facing (and thus warmer) slopes and near lakes, rivers, bays and islands, all now submerged.

That report came out in 2013, and led directly to a US\$600,000 grant from BOEM to seek out evidence for the predictions about prehistoric environments. On a series of cruises off California and Oregon over the next three years, researchers will use a variety of sonar instruments to survey the ocean floor and sediments below. If they identify a possible estuary, beach or other ancient shoreline

feature, they will take core samples and carbon date biological material from the various layers of sediment to confirm the find.

Principal investigator Todd Braje, an anthropologist at San Diego State University in California, is trying to expand the project by encouraging the US National Oceanic and Atmospheric Administration and other potential funders to add more cruises. But even at its current size, he says, "This is probably the biggest effort to identify submerged sites along the Pacific coast."

The investment may be big, but Braje is trying to keep expectations modest. He insists that the goal is to learn how to identify environments in which humans might have camped or settled up to 20,000 years ago; the team is not expecting to find the remains of any settlements, and certainly not ones older than those of the Clovis settlers. "The idea that we are going to hit on a 15,000-year-old site that is underwater is probably unrealistic in the near future," says Braje. "You get to those first migrants into the New World and the archaeological footprint they left is very small."

The project will build on Davis's model of submerged environments, using coring and imaging to test whether his projections actually lead them to the right sorts of sites. Davis is a co-principal investigator and will join the Oregon cruises next year.

In the meantime, he is digging in Idaho. It is near the end of the field season, and he and his crew are working on their day off to finish as much as possible. He has bribed them with gourmet cheese, and he lays it out with no fewer than five cheese knives. Combined with the trowels, brushes, scrapers and spoons used by the crew, the site is bristling with tools.

Given all the difficulties of this work, those involved in investigating the ocean-migration hypothesis stress that expectations should remain modest for many years as researchers improve their search methods. If the theory is correct, the first definitive older-than-Clovis find along the coast — the green light for the theory that everyone seems to be hoping for — could still be far off. "It could happen this summer, next summer, it could be ten years," says Erlandson.

Or it could happen right now in the sweltering pit at Cooper's Ferry, with the very next scrape of a trowel. ■

**Emma Marris** is a freelance writer in Klamath Falls, Oregon.

1. Davis, L. G., Nyers, A. J. & Willis, S. C. *Am. Antiquity* **79**, 596–632 (2014).
2. McLaren, D. *et al. Quat. Sci. Rev.* **97**, 148–169 (2014).
3. Fedje, D. W. & Josenhans, H. *Geology* **28**, 99–102 (2000).
4. ICF International, Southeastern Archaeological Research & Davis Geoarchaeological Research. *Inventory and Analysis of Coastal and Submerged Archaeological Site Occurrence on the Pacific Outer Continental Shelf* (BOEM, 2013).

# COMMENT



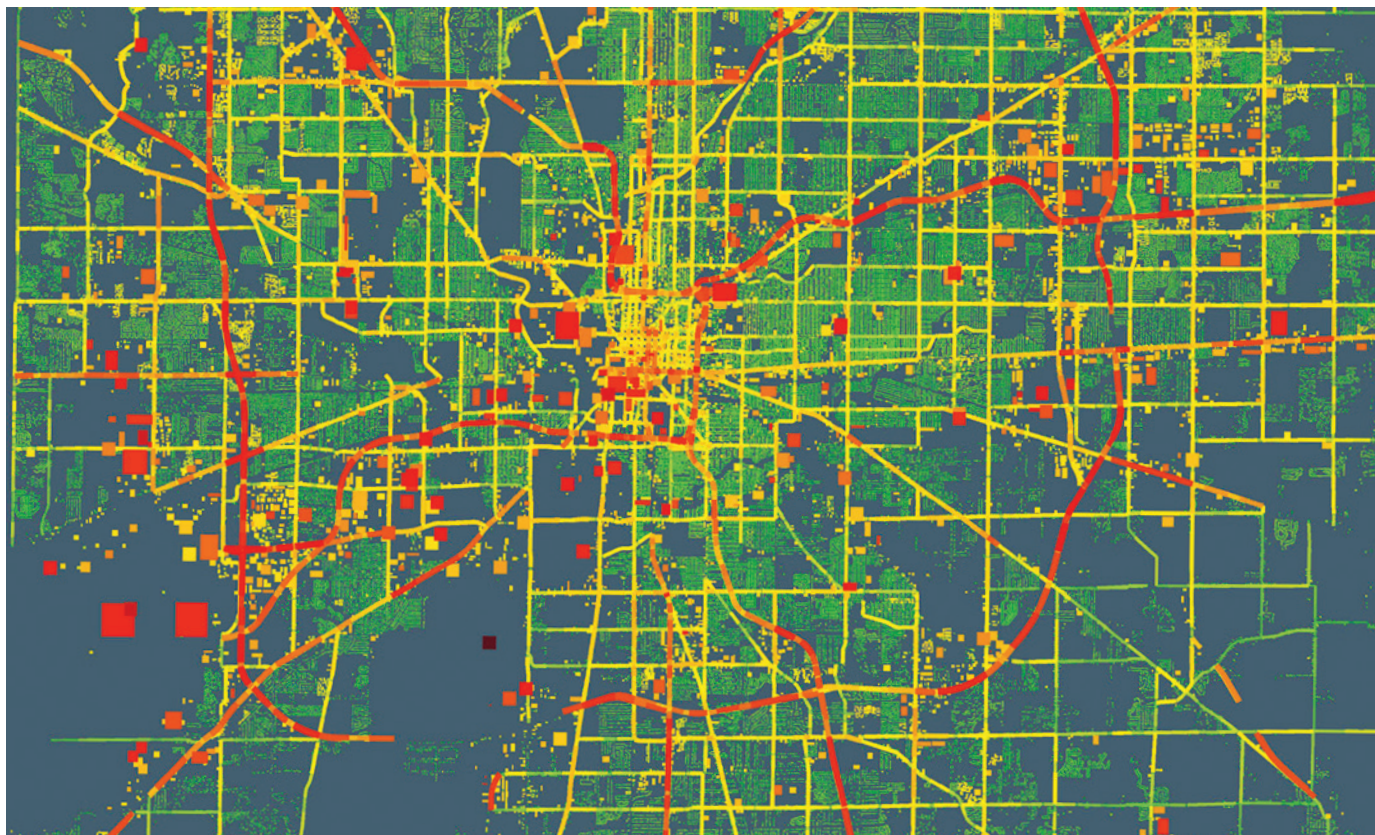
**FUNDING** Women get fair share of UK social-science grants **p.181**

**GENETICS** Recent research gets short shrift in part two of Dawkins's autobiography **p.184**

**FILM** Cartoonist creator of *The PhD Movie 2* on the scramble for grants and jobs **p.186**

**OBITUARY** Oliver Sacks, chronicler of patients' experience, remembered **p.188**

MICHEL ABUL-MASSIH & BEDRICH BENES/PURDUE UNIV.



A visualization of carbon dioxide emissions data from Marion County, Indiana, shows that large buildings and main roads (red areas) emit the most.

## Track urban emissions on a human scale

Cities need to understand and manage their carbon footprint at the level of streets, buildings and communities, urge **Kevin Robert Gurney** and colleagues.

Cities are taking steps to combat climate change, given the scant progress made by international treaty negotiations. Los Angeles, California, home to around 4 million people, has one of the most ambitious targets: to reduce greenhouse-gas emissions by 35% below 1990 levels by 2030. The city has calculated its carbon 'footprint' and found that road vehicles constitute 47% of total carbon dioxide emissions, and that

electricity consumption constitutes 32%<sup>1</sup>. So how should Los Angeles target its policies?

Knowing that certain roads, types of vehicle or parts of a city dominate road emissions and why people drive at specific times would tell city planners where and how to lower emissions efficiently. Improvements in traffic congestion, air quality, pedestrian conditions, and noise pollution could be aligned. But tracking emissions road by

road and building by building is beyond the capacity of most cities.

Luckily, scientists are gathering the data that city managers need — in studies that match sources of CO<sub>2</sub> and methane with atmospheric concentrations. Now the research community needs to translate this information into a form that city managers can use. Emissions data need to be merged with socio-economic information such ▶



▶ as income, property ownership or travel habits, and placed in software tools that can query policy options and weigh up costs and benefits. And scientists should help municipalities to raise awareness of the power of detailed emissions data in tailoring climate and development policies.

### CARBON HOTSPOTS

Cities account for more than 70% of global fossil-fuel CO<sub>2</sub> emissions, the main driver of climate change. If the top 50 emitting cities were counted as one country, that 'nation' would rank third in emissions behind China and the United States<sup>2</sup>. Urban areas are set to triple globally by 2030 (ref. 3).

Much of this emitting landscape falls within the control of mayors, city planners, businesses and community groups that have responsibility for residents' health and well-being. A 2014 survey lists 228 global cities — representing nearly half a billion people — that have pledged reductions equivalent to 454 megatonnes of CO<sub>2</sub> per year by 2020 (see [go.nature.com/inaxr4](http://go.nature.com/inaxr4)). Shenzhen in China, for example, aims to put an extra 35,000 electric vehicles on the road by the end of 2015. The German city of Munich aims to produce enough green electricity by 2025 to meet all its power requirements.

Yet such pledges account for only about 3% of global urban emissions and less than 1% of total global emissions projected for 2020 (ref. 4). Rich cities dominate these pledges, yet low- and middle-income countries are experiencing the greatest urban growth.

Slashing emissions requires mapping them on finer scales of space and time that reflect the human dimensions at which carbon is emitted: by individual buildings, vehicles, parks, factories and power plants. These should be tracked at least yearly. Such granular estimates are needed for

several reasons: to verify emissions rates; to confirm progress towards reduction and support carbon trading, permits or taxation<sup>5</sup>; to enable more-targeted and financially efficient decisions about mitigation options<sup>6</sup>; and to identify and fix unintentional releases from, for example, leaking gas pipes or malfunctioning methane-capture equipment in landfills.

Cities already approach air-quality improvement, regional development, transport planning and waste disposal on a house or road scale. Adding low-carbon policies to these efforts could benefit them all. For example, reducing traffic congestion would lower air pollution and traffic accidents and improve commutes. And targeting residents' immediate needs widens public acceptance.

### THE PROBLEMS

Although methods to account for community-scale emissions have been designed by non-profit organizations such as the World Business Council for Sustainable Development and the World Resources Institute (see [go.nature.com/q7wjeb](http://go.nature.com/q7wjeb)), most cities lack independent, comprehensive and comparable sources of data. The expertise and staff required to build this information are costly. Transparency of data and methods is also crucial to enable verification by third parties and to build trust.

Scientists are starting to meet these challenges. In the past five years, 'bottom-up' estimations of carbon emissions from fuel reporting, traffic data, building information and human activity are being merged with 'top-down' atmospheric measurements over cities of CO<sub>2</sub>, methane and <sup>14</sup>CO<sub>2</sub> — an isotope of CO<sub>2</sub> that reflects fuel combustion<sup>7</sup>. Such efforts began in the late 2000s in Paris and in the US cities of Indianapolis, Boston,

Salt Lake City and Los Angeles; more are planned for São Paulo, Brazil, and cities in Australia, China, the United Kingdom and Canada. These studies cost millions of dollars, and involve at least a dozen monitoring sites and analysis of remotely sensed data and modelling efforts. Many of these data sets are now publicly available.

Links between ground-based and satellite remote sensing are improving. For example, Japan's Greenhouse Gases Observing Satellite (GOSAT) has shown<sup>8</sup> that space-borne CO<sub>2</sub> measurements can constrain

**"The measurement, monitoring and modelling of urban carbon flows is a global need."**

the 'domes' of the gas that lie above cities. This work will continue with NASA's Orbiting Carbon Observatory 2 (OCO-2), which launched in July 2014.

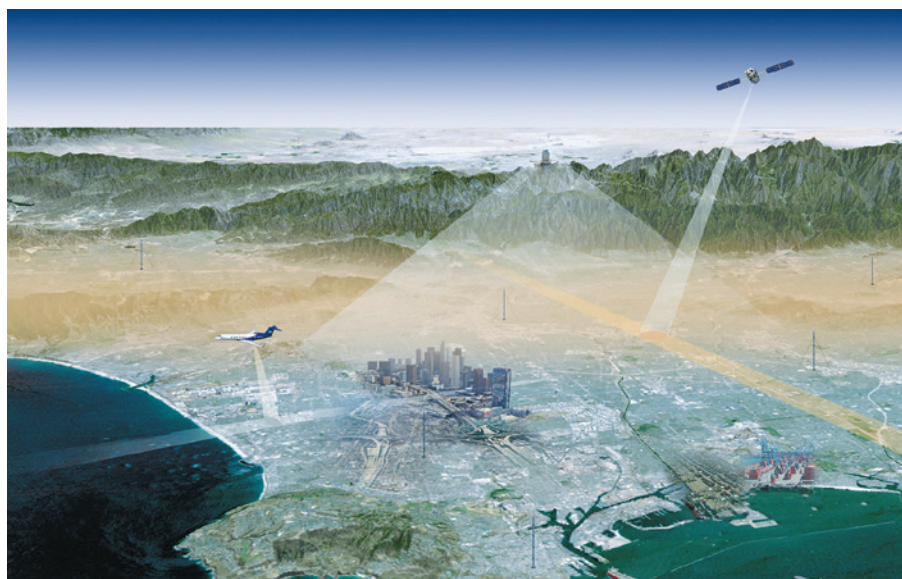
Future space missions (such as OCO-3, planned for 2018) will have a 'city mode' that will monitor urban areas and power plants monthly. The European Space Agency's Sentinel-5 mission (to launch in 2016) should provide near-global measurements of large methane emitters on the urban scale every few days or weeks. Work is also under way to characterize infrastructure in high-resolution images. Complemented by ground-based information such as traffic data from mobile phones, this could reveal, for example, which types of building or location account for disproportionate urban emissions and why.

Much needs to be known to make the science suitable for policymakers and planners. For example, what level of granularity and accuracy is most useful? How many atmospheric monitoring stations are sufficient to calibrate or anchor emissions inventories? How does this scale with city size or types of emission (road versus industrial, say)?

Existing information systems are cumbersome, and although good at quantifying emissions, they are unable to explain the roots and controls of carbon flows. Researchers need to understand the relationships between urban carbon fluxes and the social norms, technology, economics and institutional constraints that drive emissions. This is especially important in low- and middle-income countries.

### INTERNATIONAL COLLABORATION

More collaboration among disciplines is needed. For example, engineers have modelled how emissions change when transit systems or compact urban development strategies are introduced. But technological and infrastructure changes are rarely modelled within socio-ecological



The Los Angeles Megacities Carbon Project will measure urban carbon levels from the ground and sky.

NASA/JPL-CALTECH

systems<sup>9</sup>. Social scientists are examining the connections between wealth, population size or density and carbon emissions<sup>10</sup>, but not within realistic, economically constrained, engineered landscapes.

Translating urban carbon science into solutions requires two key steps. First, it must become 'operational'. Like weather stations, data and forecasting, the measurement, monitoring and modelling of urban carbon flows is a global need that is best accomplished collectively. This requires long-term collaborative funding and institutional support beyond the typical three-year research-grant cycle.

Second, an independent intergovernmental centre (with regional representation) is needed to ensure standardization and priority. This could be funded jointly by governments, foundations and intergovernmental institutions. Such an 'urban carbon solutions centre' must generate practical results, tools and carbon-mitigation options with the involvement of community groups, mayoral staff and energy providers. Cities could pay the solutions centre to provide information tailored to their locale. Some work could be undertaken by the private sector.

With detailed knowledge of carbon flows, cities might succeed in reducing global emissions where nations have failed. ■

**Kevin Robert Gurney** is associate professor in the School of Life Sciences and the Global Institute of Sustainability, Arizona State University, Tempe, Arizona, USA. **Paty Romero-Lankao, Karen C. Seto, Lucy R. Hutyra, Riley Duren, Christopher Kennedy, Nancy B. Grimm, Jim R. Ehleringer, Peter Marcotullio, Sara Hughes, Stephanie Pincetl, Mikhail V. Chester, Daniel M. Runfola, Johannes J. Feddema, Joshua Sperling,**  
e-mail: kevin.gurney@asu.edu

1. City of Los Angeles. *ClimateLA: Municipal Program Implementing the GreenLA Climate Action Plan* (City of Los Angeles, 2008).
2. World Bank. *Cities and Climate Change: An Urgent Agenda* (Int. Bank for Reconstr. and Dev. & World Bank, 2010).
3. Seto, K. C., Güneralp, B. & Hutyra, L. R. *Proc. Natl Acad. Sci. USA* **109**, 16083–16088 (2012).
4. Erickson, P. & Tempest, K. *Advancing Climate Ambition: How City-Scale Actions Can Contribute to Global Climate Goals* (Stockholm Environ. Inst., 2014).
5. McKain, K. et al. *Proc. Natl Acad. Sci. USA* **109**, 8423–8428 (2012).
6. Gurney, K. R. *Eos* **94**, 199–200 (2013).
7. Hutyra, L. R. et al. *Earth's Future* **2**, 473–495 (2014).
8. Kort, E. A., Frankenberg, C., Miller, C. E. & Oda, T. *Geophys. Res. Lett.* **39**, L17806 (2012).
9. Chester, M. V. et al. *Earth's Future* **2**, 533–547 (2014).
10. Marcotullio, P. J. et al. *Earth's Future* **2**, 496–514 (2014).

Full author affiliations accompany this article online at [go.nature.com/vpqnf3](http://go.nature.com/vpqnf3)



## Women are funded more fairly in social science

UK data hold lessons for how to close the gender gap in bioscience grant applications, success and size, argue **Paul Boyle** and colleagues.

**D**espite the increasing commentary and debate on gender disparities in science<sup>1,2</sup>, equality will not be achieved without proactive support from key institutions.

One of the key drivers of academic inequality is the receipt of competitive grant funding. In the biomedical sciences, women get smaller grants than men in the United States<sup>3</sup> and the United Kingdom<sup>4</sup>.

**► NATURE.COM**  
For Nature's special issue on women in science, see: [nature.com/women](http://nature.com/women)

Similarly, figures from the European Research Council (ERC) for 2007–13 show that women make only one-quarter of grant applications, and they receive just one-fifth of awards. This pattern is evident at different rates across disciplinary domains: in the physical sciences and engineering, women submit 17% of grant applications and receive 15%; in the life sciences, 30% and 21%; and in the social sciences and humanities, 36% and 31% (see [go.nature.com/nqfvc3](http://go.nature.com/nqfvc3)).

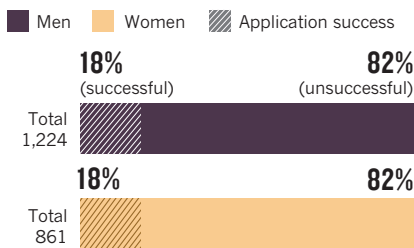
We find that UK social-science funding does not show such gender bias. When ►



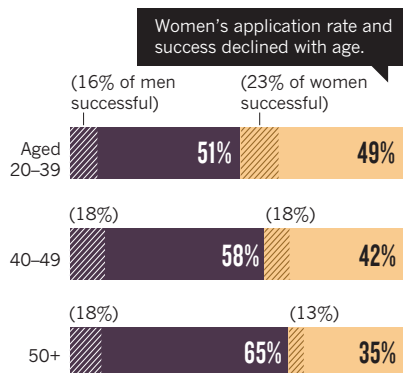
## CHECKING THE BALANCE

Data from the United Kingdom's main social-science funding body show little difference between female and male social scientists in application rate, success rate and grant size.

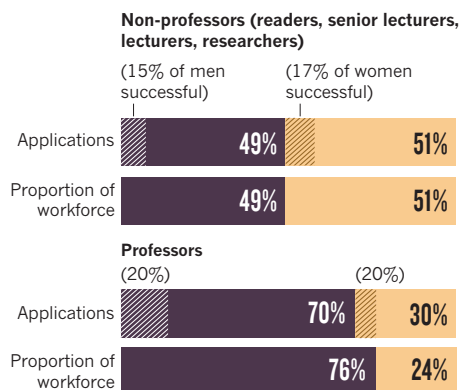
### 1 Overall grant applications and success



### 2 Grant applications with age

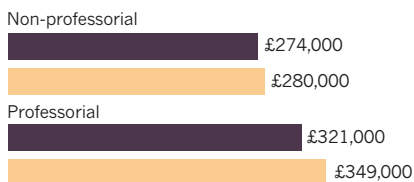


### 3 Grant applications and awards by professional grade

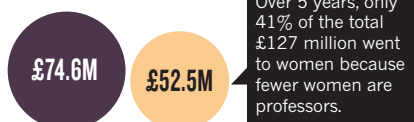


### 4 Amounts awarded

#### Median size of awarded grants



#### Total budget split



► academic position is accounted for, in the data we present here there is little difference between female and male social scientists in application rate, success rate and grant size. We discuss some lessons that these results may hold for the biomedical sciences.

### WHO GOT WHAT?

We considered applications to the UK Economic and Social Research Council (ESRC) Research Grants 'open call' scheme between 2008 and 2013 (one of us, P.B., was chief executive of the ESRC from 2010 to 2014). We examined whether women and men submitted a similar number of grants, and their respective success rates and sizes of awarded grants. We used the UK government's Higher Education Statistics Agency for data on numbers of men and women in social-science academic jobs in the United Kingdom. The results described here are underpinned by robust multivariate analyses.

We found that women were less likely than men to apply for grant funding (making up 41% of applications), even though there were only slightly fewer women (48%) than men in social-science academic posts. But women and men were equally successful in winning ESRC grants (18% success rate for both; see 'Checking the balance').

Women's application rates and funding success declined with age; men in all age groups had a similar success rate. Women under the age of 40 applied for as many grants as men in that age group, and were more successful. Women over 50 applied for fewer grants and were less successful than men in the same age group.

Comparing applications and success by academic position, we find that it is the smaller number of female professors that accounts for the overall difference in grant applications between men and women, and the greater success of older men. Male and female professors were equally successful, and women at lower grades were slightly more successful than men at the equivalent grade. Indeed, female professors were more likely to bid than their male counterparts: women made 30% of professorial applications, even though nationally only 24% of professorial posts were held by women.

The median amount awarded was not significantly different for women and men. On average, female professors won slightly larger amounts than male professors. However, over the five-year period, 59% of the £127 million (US\$198 million) allocated went to men, because fewer applications were received from women (recall that, accounting for academic position, those who applied for grants were equally successful).

The analyses we present are based on standard, routinely available data. Thus,

several caveats should be borne in mind. These include: the difficulties of combining different data sources on staff eligible to apply for awards and actual applications; and a lack of further detailed information on applications, such as the number and gender of co-applicants.

Even so, there are clearly disciplinary differences in women's funding engagement and success. A comparison of figures from the UK Medical Research Council

**"Significant change is unlikely, without some bold re-structuring."**

(MRC) and the ESRC shows that although the proportion of women in biomedical and social-science disciplines is similar (43% and 45%,

respectively), the proportion of female grant applicants in 2012–13 was 27% at the MRC and 42% at the ESRC (see [go.nature.com/pesa2z](http://go.nature.com/pesa2z)). Furthermore, ESRC grants secured by female social scientists are of comparable size to those awarded to men. By contrast, at the Wellcome Trust<sup>4</sup>, a major UK biomedical-research charity, awards between 2000 and 2008 were on average around £44,500 (around 15%) bigger for men than for women.

### DISRUPTING HIERARCHIES

Whether these differences are a result of endemic discriminatory practices that discourage women from applying for awards — and for larger ones — in biomedical disciplines should be the focus of intense scrutiny.

It is interesting to consider why women may be better served in the social sciences. The positive consequences of higher levels of female representation in social-science disciplines include a move away from 'conventional gender expectations'<sup>5</sup> that align with hierarchical, individualistic and competitive behaviours. Social scientists have long been engaged with feminist research-management practices, with the guiding principles of consultation, collaboration and social equality, which have disrupted male hierarchies<sup>6</sup>. Critiques of knowledge creation that exclude women as both researchers and participants have ensured that men in the social sciences have long been aware of the ingrained, institutionalized male culture of universities<sup>7</sup> — an awareness that may be taking longer to permeate the science, technology, engineering and mathematics (STEM) disciplines.

Even so, the lack of women in professorial positions means that 59% of the total funds disbursed by ESRC between 2008 and 2013 in this study was allocated to men. Young female social scientists of today submit similar numbers of ESRC applications as equivalent men, are as successful and receive grants

## BOLD ACTION

## Ten ways institutions must close the gender gap

## Steps for funding agencies worldwide

- Commit to ambitious expectations for gender performance that link to eligibility for receiving awards, following the lead of the National Institute for Health Research.
- Introduce targets for minimum gender representation on funding panels.
- Train selection panels on gender-equality issues, including unconscious bias.
- Submit data annually to independent scrutiny of gender differences in applications, success rates and award sizes.

- Publish figures to allow cross-agency and cross-national comparison by discipline.

## Steps for universities worldwide

- Publish gender breakdowns in key areas including promotions, appointments and rewards in a consistent way, allowing for cross-institution comparison; such transparency would allow prospective employees and students to assess the institutional culture.

- Embed gender-equality issues in work practice. Become beacons of good practice for public-sector and private employers.
- Support women's career progression through the ongoing development of promotion criteria that focus on quality rather than quantity.
- Engage men in championing gender equality. Commit to the principles and uptake of shared parental leave.
- Celebrate women's achievements equally in a public way.

of comparable size; it is quite possible that they will maintain this success as they age. Yet this is unlikely to transpire if women cannot access the more senior positions that men have dominated. And as these women rise through the ranks they will not experience the same work-life balance as men, the same child or parental care responsibilities, or the same cultural attitudes to the importance of their labour. Consequently, they will be more likely to have part-time or fixed-term contracts and to take career breaks.

In other words, young women perform well today, but they will continue to match men only if structural changes are implemented within universities and funding agencies (see 'Ten ways institutions must close the gender gap'). If the current pace of

change continues, it will take 39 years for women to be represented equally among the UK professoriate — and this is likely to vary widely by discipline (see [go.nature.com/gwihpt](http://go.nature.com/gwihpt)).

## SIGNS OF IMPROVEMENT

It would be wrong to assume that nothing is being done. All seven UK research-funding councils, through Research Councils UK, have published expectations for themselves and for institutions in receipt of their funding, and these statements include an ongoing commitment to promoting cultural change in relation to equality and diversity. The Athena SWAN initiative provides a framework for addressing gender imbalances in biomedicine and has catalysed action — particularly since the

National Institute for Health Research made attainment of an Athena SWAN silver award a requirement for certain large-scale funding.

Remarkable progress has been made elsewhere,

most notably in Nordic countries. In Finland, for example, equality legislation introduced 20 years ago requires a minimum representation of 40% of either gender on any committee responsible for public spending, including research funders. Although controversial, even among some ardent proponents of gender equality, the rule has resulted in substantial change. By 2010, women made up 50% of the board of the Academy of Finland and of the country's scientific committees.

Despite such signs of improvement, gender inequality along the science-career trajectory continues to be pervasive. Men earn more than women<sup>8</sup>; academics who are mothers are less likely to be promoted and have lower salaries than women who do not have children<sup>9</sup>; and

'Brian' is more likely to be hired than 'Karen' as a professor, even if they have identical applications<sup>10</sup>. Consequently, there are fewer women in senior professorial, administrative and university-president roles. Although women make up 47% of non-professorial higher-education positions in the United Kingdom, they account for less than 20% of professorial appointments.

Significant change is unlikely, without some bold re-structuring. Bringing together funding agencies and a consortium of prominent universities who have shown commitment to these issues to develop coordinated approaches could have a significant impact. Organizations such as Science Europe and the Global Research Council, which have already committed to helping to reduce gender inequalities in science, should lead the way. ■

**Paul J. Boyle** is president and vice-chancellor of the University of Leicester, UK. From 2010 to 2014, he was chief executive of the United Kingdom's Economic and Social Research Council. **Lucy K. Smith** is a senior research fellow, **Nicola J. Cooper** is professor of health-care evaluation research, and **Kate S. Williams** is a senior research fellow in the Department of Health Sciences, University of Leicester, UK. **Henrietta O'Connor** is professor in the Department of Sociology, University of Leicester, UK.

1. *Nature* **522**, 255 (2015).
2. Shen, H. *Nature* **495**, 22–24 (2013).
3. Reineke Pohlhaus, J. et al. *Acad. Med.* **86**, 759–767 (2011).
4. Bedi, G., Van Dam, N. T. & Munafo, M. *Lancet* **380**, 474 (2012).
5. Glass, J. L. et al. *Soc. Forces* **92**, 723–756 (2013).
6. Mauthner, N. S. & Edwards, R. *Gender Work Organ.* **17**, 481–502 (2010).
7. Morgan, D. in *Doing Feminist Research* (ed. Roberts, H.) Ch. 4 (Routledge and Kegan Paul, 1981).
8. Moss-Racusin, C. A. et al. *Proc. Natl Acad. Sci. USA* **109**, 16474–16479 (2012).
9. Correll, S. J., Benard, S. & Paik, I. *Am. J. Sociol.* **112**, 1297–1339 (2007).
10. Steinpreis, R. E., Anders, K. A. & Ritzke, D. *Sex Roles* **41**, 509–528 (1999).



ILLUSTRATION BY BELLE MELLOR





Richard Dawkins, pictured at home in 2010, popularized a gene-based view of evolutionary biology.

## GENETICS

# Dawkins, redux

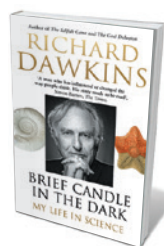
**Nathaniel Comfort** takes issue with the second instalment of the evolutionary biologist's autobiography.

In 1976, Richard Dawkins, then a 35-year-old Oxford lecturer in animal behaviour, published his first book, *The Selfish Gene* (Oxford University Press). Distilling a body of recent population-genetics research — notably that of W. D. Hamilton — it argued that genes, not organisms, were the targets of natural selection. An organism, Dawkins wrote, was simply a gene's way of replicating itself.

The book was a surprise best-seller. Along with E. O. Wilson's *Sociobiology* (Harvard University Press, 1975), it helped to spark a new nature–nurture debate that pitted sociobiologists against socialist biologists. Notable among the latter were the palaeontologist Stephen Jay Gould and the population geneticist Richard Lewontin, who accused the sociobiologists of rationalizing social evils such as racism and infidelity as genetically hard-wired, evolutionarily programmed. Yet

Dawkins's fiercely reductionist, materialist world view exuded a transgressive sexiness, and his suave, swaggering prose appealed to many readers, lay and professional.

Dawkins's greatest gift has been as a lyricist. With terms such as selfish genes, memes and the extended phenotype, he has provided much of the vocabulary of modern evolutionary biology. He has published a sackful of books laying out the evidence for evolution, against design in nature, and for natural selection as the only mechanism of adaptive evolution. A skilled and popular lecturer,



**Brief Candle in the Dark: My Life in Science**  
RICHARD DAWKINS  
Bantam: 2015.

he also discovered a taste for the camera, hosting numerous television documentaries.

In the early 2000s, he saltated from popularizer into evangelist. His 2006 book *The God Delusion* (Bantam) was an ecclesiophobic diatribe, published around the same time as Christopher Hitchens's *God Is Not Great* (Twelve, 2007) and similar books by Daniel Dennett and Sam Harris. The gospels of Christopher, Daniel, Sam and Richard form the scripture of the 'new atheism', a fundamentalist sect that has mounted a scientific crusade against all religion.

Now, *Brief Candle in the Dark* revisits Dawkins's career since *The Selfish Gene*. Its predecessor, *An Appetite for Wonder* (Bantam, 2013), was a memoir of a young upper-class Englishman becoming a scientist, replete with African adventures, British public schools and Oxonian traditions. Some reviewers wondered whether the sequel would have more heft and focus, reflection and introspection. At 450 pages, it is certainly heftier.

Dawkins has organized *Brief Candle* thematically, making it less memoir than annotated catalogue. The first few chapters are a scattershot record of his duties as an Oxford don, a rare field trip and the Royal Institution Christmas lectures. The next few devolve into a series of lists: his books, his debates, his television appearances.

Finally, he abandons the memoir format to do what he does best: write about science. The book concludes with a mammoth 120-page chapter recapitulating the ontogeny of his thought. Like Francis Galton, the hereditarian Victorian biostatistics pioneer, Dawkins has a quantitative turn of mind, but is better at algorithms than theorems. So indeed is life itself, which is why biology has so few laws.

Much of Dawkins's research has been *in silico*, writing programs for evolutionary simulations. In his simulations, life is utterly determined by genes, which specify developmental rules and fixed traits such as colour. The more lifelike his digital animals ('biomorphs') become, the more persuaded he is that real genes work in roughly the same way. Dawkins's critics accuse him of genetic determinism. This synopsis of his work shows that his life virtually depends on it.

A curious stasis underlies Dawkins's thought. His biomorphs are grounded in 1970s assumptions. Back then, with rare exceptions, each gene specified a protein and each protein was specified by a gene. The genome was a linear text — a parts list or computer program for making an organism — insulated from the environment, with the coding regions interspersed with 'junk'.

Today's genome is much more than a script: it is a dynamic, three-dimensional structure, highly responsive to its environment and almost fractally modular. Genes may be fragmentary, with far-flung chunks of DNA sequence mixed and matched in bewildering

combinatorial arrays. A universe of regulatory and modulatory elements hides in the erstwhile junk. Genes cooperate, evolving together as units to produce traits. Many researchers continue to find selfish DNA a productive idea, but taking the longer view, the selfish gene per se is looking increasingly like a twentieth-century construct.

Dawkins's synopsis shows that he has not adapted to this view. He nods at cooperation among genes, but assimilates it as a kind of selfishness. The microbiome and the 3D genome go unnoticed. Epigenetics is an “interesting, if rather rare, phenomenon” enjoying its “fifteen minutes of pop science voguery”, which it has been doing since at least 2009, when Dawkins made the same claim in *The Greatest Show on Earth* (Transworld). Dawkins adheres to a deterministic language of “genes for” traits. As I and other historians have shown, such hereditarianism plays into the hands of the self-styled race realists (N. Comfort *Nature* 513, 306–307; 2014).

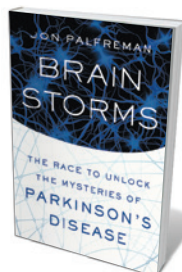
His writing can still sparkle. He excels at capturing the scenes behind a scene, deftly explaining a scientific principle, capping a story with an amusing anecdote. His tale of palaeoanthropologist Richard Leakey hauling his legs (amputated after a plane crash) to Kenya in his hand luggage for burial is funny and touching. Dawkins also makes an important case for the “poetic” side of science, arguing that the imperative to justify research in terms of potential medical or financial benefits bleeds the beauty out of it. Amen.

At such moments, one feels transported to a tweedy evening at Oxford, pouring the sherry as a charming senior faculty member holds court. But too often, the professor rambles. He quotes friends’ and colleagues’ tributes from dust-jackets and afterwords. He mentions the fish genus *Dawkinsia*. He repeatedly slams his late rival, Gould (“whose genius for getting things wrong matched the eloquence with which he did so”). His digressions often come off as twee and self-indulgent. Mentioning the limping family dog, Bunch, in an apt example of an acquired characteristic that cannot be inherited, he is reminded of an unfinished poem his mother wrote after Bunch died, which he prints. “If you can’t be sentimental in an autobiography, when can you?” he asks.

For a time, Dawkins was a rebellious scientific rock star. Now, his critique of religion seems cranky, and his immovably geocentric universe is parochial. *Brief Candle* is about as edgy as Sir Mick and the Rolling Stones cranking out the 3,578th rendition of ‘Brown Sugar’ — a treat for fans, but reinscribing boundaries rather than crossing them. ■

**Nathaniel Comfort** is professor of the history of medicine at Johns Hopkins University in Baltimore, Maryland. His latest book is *The Science of Human Perfection*. Twitter: @nccomfort

## Books in brief



### Brain Storms: The Race to Unlock the Mysteries of Parkinson's

Jon Palfreman FARRAR, STRAUS & GIROUX (2015)

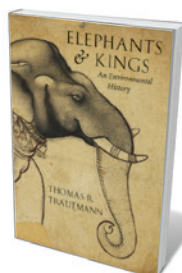
In 2011, journalist Jon Palfreman was diagnosed with Parkinson's disease. The progressive neurodegenerative condition, characterized by tremors and muscular rigidity, affects 7 million people worldwide. In this lucid overview, Palfreman interlaces the history of research into the disease — linked, like Alzheimer's, to a rogue protein — with therapeutic approaches from deep brain stimulation to the drug L-DOPA. Extraordinary case studies abound, such as that of a man who can ride a bicycle but not walk, and dancer Pamela Quinn, who has devised workarounds that ‘trick’ the body into movement.



### Places of the Heart: The Psychogeography of Everyday Life

Colin Ellard BELLEVUE LITERARY (2015)

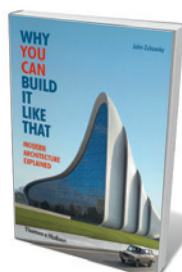
Why would a street evoke unease, or a shopping centre the desire to spend? Psychologist Colin Ellard explores the intersection of neuroscience and urban design for answers. Meshing recent findings with thoughtful appraisals of their implications, Ellard looks at spaces and the awe, lust, boredom, affection or anxiety that they trigger. He is richly insightful, particularly on digital encroachments into the experience of place: can augmented-reality gear ever vie with the hair-prickling thrill of being there? Ellard argues that virtual immersion could take a “metaphysical toll”; it is hard not to agree.



### Elephants and Kings: An Environmental History

Thomas R. Trautmann UNIVERSITY OF CHICAGO PRESS (2015)

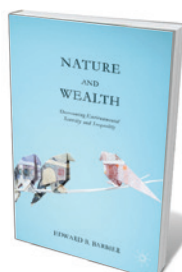
The intelligence, majestic presence and physical prowess of the Asian elephant was not lost on India's monarchs. As historian Thomas Trautmann shows in this scholarly environmental history, the beast's usefulness in warfare and its prodigious dietary needs ensured royal protection for swathes of forest in ancient India, where the wild animals were captured for specialized training. That the country still has 30,000 elephants is a testament to their enduring place in the collective imagination; but as Trautmann argues, India's surviving patchwork of 31 elephant reserves may not sustain them.



### Why You Can Build It Like That: Modern Architecture Explained

John Zukowsky THAMES & HUDSON (2015)

From the squat circularity of New York City's Guggenheim Museum to Abu Dhabi's swooning, tornado-shaped Capital Gate skyscraper, extreme architecture is here to stay. This illustrated roll call by architectural historian John Zukowsky zips through 100 “iconic and iconoclastic” structures of the past 50 years — shapely, hideous or energizingly weird. Norman Foster's Spaceport America in New Mexico, for instance, resembles a giant horseshoe crab in thin-shelled concrete, whereas Myron Goldsmith's McMath-Pierce Solar Telescope in Arizona is a minimalist ode to the right angle.



### Nature and Wealth: Overcoming Environmental Scarcity and Inequality

Edward B. Barbier PALGRAVE MACMILLAN (2015)

In this cogent analysis, economist Edward Barbier reveals an economic landscape of degraded environments and social inequality. The culprit, he argues, is a structural imbalance in which natural resources are overexploited and human capital is undersupplied. Examining current challenges such as ecological scarcity, he concludes that a strategy to rebalance natural and human capital is the way forward, however difficult. *Barbara Kiser*



# Correspondence

## Outdated listing puts species at risk

Loopholes could allow illegal wildlife traders and hunters in China to evade prosecution or to receive reduced sentencing. The problem stems from China's Protected Species List (PSL): this has not been updated since it was implemented in 1989, resulting in incongruity with newer taxonomy.

Appendices I and II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and the International Union for Conservation of Nature (IUCN) use taxonomic classifications based on recent revisions to geographical distributions and phylogenetic relationships. Some species that were formerly listed as exotic to China under CITES have had their listing changed to native. But because the PSL has not been revised accordingly, the endemic status of such species is not recognized in law.

For example, the Chinese pangolin (*Manis pentadactyla*) is on the PSL as a native species. The endangered Malayan and Indian pangolins (*Manis javanica* and *Manis crassicaudata*) receive protection as exotic animals under CITES II, but have been endemic to the country since at least 2000. In our view, these species should be added to the PSL immediately to ensure that they have the same protected status as the Chinese pangolin in national legislation. The taxonomic status of leaf monkeys (*Trachypithecus* spp.) and the Burmese python (*Python bivittatus*) have also become inconsistent with the PSL, leading to similar risks for these species.

As a result, illegal traders can claim that these animals with biogeographical or name revisions are not on the PSL, even though they may be endangered. To alleviate trans-border inconsistencies and aid enforcement, this naming

inconsistency issue requires that all 181 signatory nations to CITES adopt unambiguous standardized and internationally coherent naming policies, following the IUCN Red List and CITES Species+ ([www.speciesplus.net](http://www.speciesplus.net)).

**Zhao-Min Zhou\*** *Yunnan Public Security Bureau for Forests, Kunming, Yunnan, China.*  
[zhouzm81@gmail.com](mailto:zhouzm81@gmail.com)

*\*On behalf of 6 correspondents (see [go.nature.com/hubzzy](http://go.nature.com/hubzzy) for full list).*

## Physicists' report on EU green electricity

The European Physical Society has released a report on European Union (EU) plans for sustainable production of green electricity in the context of today's global energy and climate challenges (see [go.nature.com/2blxp9](http://go.nature.com/2blxp9)). The report advises Europe to develop a common energy policy that could act as a template for other regions.

It points out that Europe's contribution to global greenhouse-gas emissions is relatively low, so producing electricity without fossil fuels would cut global emissions by a mere 3–4%. Any plans for worldwide green electricity structures would need to address problems such as intermittency and storage, and the need for backup systems and large, high-capacity electrical grids.

The report suggests that energy targets should be scientifically justifiable and adjusted to be more realistic. This would reduce the cost of enforcing the targets through regulation and encourage competition in EU industry. A common policy is needed for implementing those regulations, the report emphasizes.

It also recommends that the public should have access to scientific information on energy issues rather than to simplified plans and projections, and it

urges Europe to continue to lead the way in cutting greenhouse-gas emissions.

**Jozef Ongena** *Laboratory for Plasma Physics, Royal Military Academy, Brussels, Belgium.*

**Christophe Rossel** *European Physical Society, Mulhouse, France.*

[j.ongena@fz-juelich.de](mailto:j.ongena@fz-juelich.de)

## Laboratory seawater studies are justified

In our view, your report 'Seawater studies come up short' (*Nature* 524, 18–19; 2015) fails to capture the nuances of the survey results you discuss (see C. E. Cornwall and C. L. Hurd *ICES J. Mar. Sci.* <http://doi.org/68g>; 2015).

Researchers aim to follow the 2010 'Guide to best practices for ocean acidification research and data reporting' ([go.nature.com/sp5kgn](http://go.nature.com/sp5kgn)) as they strive to understand how marine organisms are likely to respond to the falling pH of the world's oceans, caused by increased carbon dioxide concentrations. The logistical constraints of testing the effects of seawater acidification on marine life in the laboratory are considerable. However, experiments must be replicated while complying with the requirements for manipulating and monitoring seawater carbonate chemistry.

The paper, co-authored by two of us, does not conclude from its meta-analysis of such manipulation experiments that all these studies "come up short". Rather, it uses them to highlight the importance and challenges of proper experimental design for such testing. The examples of experimental pitfalls it cites are intended not as criticisms, but to guide future efforts.

Together with palaeo-record investigations, modelling studies and natural and manipulated field experiments, we believe that laboratory experiments are crucial to the mechanistic understanding and prediction of

ocean-acidification impacts.

**Catriona L. Hurd\*** *Institute for Marine and Antarctic Studies, University of Tasmania, Australia.*

[catriona.hurd@utas.edu.au](mailto:catriona.hurd@utas.edu.au)

*\*On behalf of 7 correspondents (see [go.nature.com/vucoxr](http://go.nature.com/vucoxr) for full list).*

## More extensive tests for e-cigarettes

We are concerned that the focus on nicotine in electronic cigarettes is causing other associated risks to be underestimated (see *Nature* 523, 267; 2015).

For example, there are no moves to evaluate the chemical hazards attributable to the non-nicotine components of e-cigarettes, such as vaporizing solvents or liquid 'flavourings' — let alone to regulate them. Neither will the latest European legislation on tobacco and related products help (see [go.nature.com/yufgfv](http://go.nature.com/yufgfv)). It does not cover liquids or devices that are branded as 'nicotine-free': however, these frequently contain nicotine and escape safety testing through erroneous labelling (see C. Hutzler *et al. Arch. Toxicol.* 88, 1295–1308; 2014).

Electronic vaping could also offer a route for using drugs such as cannabis (A. J. Budney *et al. Addiction* <http://doi.org/7cc>; 2015). With the technology set to evolve and spread rapidly, more rigorous and extensive evaluation is urgently needed.

**Frank Henkler, Andreas Luch** *German Federal Institute for Risk Assessment (BfR), Berlin, Germany.*  
[andreas.luch@bfr.bund.de](mailto:andreas.luch@bfr.bund.de)

### CONTRIBUTIONS

Correspondence may be submitted to [correspondence@nature.com](mailto:correspondence@nature.com) after consulting the author guidelines at <http://go.nature.com/cmchno>.

# Oliver Sacks

## (1933–2015)

Neurologist who made house calls.

The final stage of life, Dante wrote, is like a ship, gradually lowering its sails at the approach of the harbour. It is a serene image of destination — and eminently unfit to characterize the final decade of Oliver Sacks's life. Against a tide of diminishing health, he added four books to an already impressive catalogue: *Musophilia* (2007), *The Mind's Eye* (2010), *Hallucinations* (2012) and — only months ago — *On the Move*, a candid sequel to his childhood memoir *Uncle Tungsten* (2001). Several more books are nearly finished.

Oliver Wolf Sacks, who died in New York City on 30 August, was born in London in 1933 into a large Jewish family. His father was a general practitioner, his mother a surgeon. His aunts and uncles were inventors, chemists and physicians. He grew up with the sense that it was a family duty to be 'scientific'. In 1939, at the start of the Second World War, he was sent away to a boarding school in the English Midlands. Sacks, who would rather have been in danger with his family than safe without them, spent four miserable years there. The experience scarred him for life: "sent away" is how he put it 75 years later in the opening sentence of *On the Move*.

Reunited with his family in 1943, Sacks developed a passion for chemistry. Although he eventually chose to study medicine, contemplation of the periodic table never ceased to soothe him in times of turbulence. Sacks studied at Queen's College, University of Oxford, UK, qualifying as a physician in 1958. He left for the United States in the early 1960s and began five years of medical training, interspersed with riding motorcycles, working out in gyms, experimenting with amphetamines and lifting weights on Muscle Beach, California. When a stint in a neurochemistry lab ended with a resounding, "Sacks, you are a menace in the lab! Why don't you go and see patients — you'll do less harm", he decided to do just that. In 1965, he took up consulting at Beth Abraham Hospital in the Bronx, New York.

In its wards he encountered some 80 survivors of the 'sleepy sickness' pandemic of the 1920s. He found them to be frozen, mostly, in a statuesque, 'parkinsonian' state. High doses of the Parkinson's disease drug L-dopa 'awoke' them from their lethargy, but — as indicated by their vocabulary, likes, dislikes and skills — in a state of mind belonging to 40 years before and in a world that was no longer theirs. Sacks



noticed such diverse reactions from patient to patient that he adapted what was initially intended to be a conventional double-blind trial to a series of case histories, which he published in 1973 as *Awakenings*.

After reading *Awakenings*, Russian neuropsychologist Aleksandr Luria sent Sacks a letter. He praised Sacks's talent for observation and description, which reminded him of the nineteenth-century tradition of the neurological narrative. Much of what was to become vintage Sacks unfolded from this book. His work was case-oriented rather than population based, descriptive and intimate rather than detached. And he wrote books, not series of papers in neurological journals. To this he added his signature approach of making house calls. He tried to meet his 'cases' in their natural surroundings. He observed, for example, a surgeon with Tourette's syndrome while he was operating; visited Temple Grandin, a woman with autism, in her office in the animal-sciences department of Colorado State University in Fort Collins; and immersed himself in the world of deaf culture.

The case histories in *The Man who Mistook His Wife for a Hat* (1985) secured him a worldwide audience. It also helped to articulate his scientific credo. Taking inspiration from German neurologist Kurt Goldstein, Sacks came to think of neurological disorders as challenges to finding a new

equilibrium. In response to injury or disease, people go through a phase of adaptation and reorganization, often mobilizing inner resources that have previously lain dormant. According to Sacks, it is the physician's task to help patients to achieve a new order by being sensitive to these altered orientations.

As Sacks indicated in *On the Move*, growing up at a time when homosexuality was still listed as a mental disorder by the American Psychiatric Association alerted him to the sometimes detrimental consequences of psychiatric labelling. Rather than locking individuals in a 'condition', he took the upbeat perspective of pointing out the benefits over the deficits — sometimes to the point of eclipsing the original pathology. In many cases, this had a liberating effect: one may have Tourette's syndrome and still become a surgeon, or, like Grandin, have autism and have a career in science. Sacks thought in terms of neurodiversity — the idea that conditions result from normal variation — well before the term became common among those who distanced themselves from the medical perspective on autism.

Sacks saw himself as a storyteller, not a theorist. He often said that he was happy to present the case material that others could use to devise grand theories. But each story, of course, is a theory. Like Goldstein and Luria before him, he let his case histories shore up the theory of the brain as an organ that should be understood holistically, as an organism capable of plasticity and compensation. Although not the inventor of the neurological narrative, Oliver was certainly its culmination. For the coming decades, his legacy will be safe in the hearts and minds of millions of readers.

In conversation, I once brought up his numerous honorary degrees, awards and fellowships — but Oliver was quick to raise his hand to halt me, and said simply that he believed he was a good doctor. He felt that his parents recognized that he had become a careful and perceptive neurologist. Even in his eighties, being a good son was still a defining ambition of his life. ■

**Douwe Draaisma** is professor of the history of psychology at the University of Groningen in the Netherlands. He interviewed Oliver Sacks in 2005 for his book *The Nostalgia Factory* and stayed in contact with him. e-mail: d.draaisma@rug.nl

ELENA SEIBERT/CORBIS OUTLINE



## ECONOMICS

# Simple market models fail the test

An analysis of energy markets with prices that vary according to demand finds that this market design unexpectedly serves to amplify, rather than dampen, fluctuations in power use.

ALEX PENTLAND

From the dynamic pricing of electricity to congestion-based road tolls, simple market models lie hidden within much of our current thinking about government, regulation and policy. However, this sort of market thinking can easily go wrong, as exemplified by Krause *et al.*<sup>1</sup> in *Physical Review E*. The authors find that variable energy prices that are designed to adapt demand to supply, and thereby dampen fluctuations in power use, in fact amplify these fluctuations.

Variable pricing in power markets and other market models is based on what the eighteenth-century economist Adam Smith called the “invisible hand” — the idea that market competition efficiently allocates resources according to need. This idea was mathematically codified by the start of the twentieth century by working out what self-interested, rational individuals would do in response to different price, usage and market conditions, and what effect this would have on prices and demand.

Subsequent criticism of the ‘rational’ part of this model has given rise to what is now known as behavioural economics. However, human limitations on rationality — which simply means that individuals know their goals and act to achieve them — typically only bias the market response, and do not invalidate the simple market model. In fact, there are far more serious problems concealed within the rational-individual market model. The two biggest limitations are: the focus on how the average behaviour of independent actors (often assumed to have a normal distribution) determines where the market eventually settles (reaches equilibrium); and the idea that people act independently. If we applied these assumptions to a classroom in which all the students copy answers from each other, and half the students have perfect scores but half have failed, we would declare the class a success because the average grade is a pass.

Krause *et al.* demonstrate what can go wrong when we apply this average and equilibrium thinking to a situation such as a power market, in which people’s needs and the external situation vary hour by hour and day by day, so averages and long-term behaviour do not



ANTONIO ARCOS/GETTY

**Figure 1 | Synchronization spikes.** Power markets with variable pricing are designed to shift individuals’ power consumption to low-demand times of day. But Krause and colleagues’ modelling<sup>1</sup> suggests that synchronized consumer behaviour may result in amplified spikes in demand.

capture the full picture. The energy-market model discussed by the authors is similar to that of proposed US policy to vary price by demand, with the intention that individuals will shift activities that use a lot of energy (such as air conditioning or heating) to low-demand, cheaper times of day. The goal of the policy is to smooth demand over time, allowing the power company to use more-efficient and cleaner energy sources.

The authors find that this sort of market thinking is too simplistic. They show that when there are unusual events, such as an especially hot day or a snowstorm, people’s actions become synchronized, with everyone turning up the air conditioning or the heating at the time the price normally drops (Fig. 1). As a consequence, rather than smoothing demand as expected, the market will cause huge spikes in demand that completely swamp the electricity grid, decoupling price and demand, and potentially causing power failures and even damaging the grid itself. Interestingly, this same synchronization process is thought to be the source of ‘flash crashes’ that have been observed in financial markets, in which

markets shed huge amounts of value in a few seconds as high-frequency trading algorithms become synchronized<sup>2</sup>.

In addition to issues arising from a focus on averages and long-term behaviour, the assumption that people act independently can cause other problems. Obviously, people do not act independently — they talk to one another and learn from each other. Not only is their behaviour sometimes synchronized as a response to changing external conditions, as discussed by Krause and colleagues, but also people actively make their actions more similar by trying to copy the successful actions of others. These peer-to-peer interactions drive the evolution of culture, norms and even the law. In the context of market-like situations, however, people learning from each other can lead to financial bubbles, political upheavals and health fads<sup>3,4</sup>. Such undesirable outcomes are particularly likely when there are large, rapid changes in the environment, or when the communication links between people are particularly strong and influential.

The advent of social media and the crowdsourcing of news has made it much easier

for people to learn from each other, and the pace, inclusiveness and influence of social learning and opinion change have increased dramatically. The paradoxical consequence of this technological move towards greater transparency, democratization and engagement is that fads, political turmoil and bubbles are more common than ever before. Because so many institutions are based on some sort of simple market model, the limitations of this way of thinking have become increasingly clear<sup>5,6</sup>.

What can we do about this problem of overly simple market-based policies? It took half a century for policymakers and the public to understand the simple market model's connection between price and demand. I don't think we can wait another 50 years for

people to move to a more sophisticated way of thinking that accounts for synchronization and connections between people.

Fortunately, we can simulate situations and visualize the results on today's ubiquitous digital devices. If we make computational modelling of social behaviour a standard part of policy debate, as Krause and colleagues have done, we can hope to markedly accelerate a transition towards better and more robust social policy. Scientists, policymakers and science-funding agencies should make this sort of computational social-science modelling a regular part of their portfolio. ■

Alex Pentland is at the Connection Science and Human Dynamics laboratories,

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. e-mail: pentland@mit.edu

1. Krause, S., Börries, S. & Bornholdt, S. *Phys. Rev. E* **92**, 012815 (2015).
2. Pan, W., Pentland, A. S., Cheng, R. & Emsbo-Mattingly, L. *MIT Sloan Mgmt Rev.* **54**, 54404 (2013).
3. Pentland, A. *Social Physics: How Social Networks Can Make Us Smarter* (Penguin, 2014).
4. Braha, D. *PLoS ONE* **7**, e48596 (2012).
5. Naim, M. *The End of Power: From Boardrooms to Battlefields and Churches to States, Why Being in Charge Isn't What It Used to Be* (Basic Books, 2013).
6. World Economic Forum *Outlook on the Global Agenda 2015* <http://reports.weforum.org/outlook-global-agenda-2015> (2014).

This article was published online on 2 September 2015.

## CELL BIOLOGY

# Countercurrents in lipid flow

**Two studies find that a lipid-exchange cycle mediates the enrichment of the lipid phosphatidylserine in the cell membrane compared with the membrane of an organelle called the endoplasmic reticulum, where the lipid is produced.**

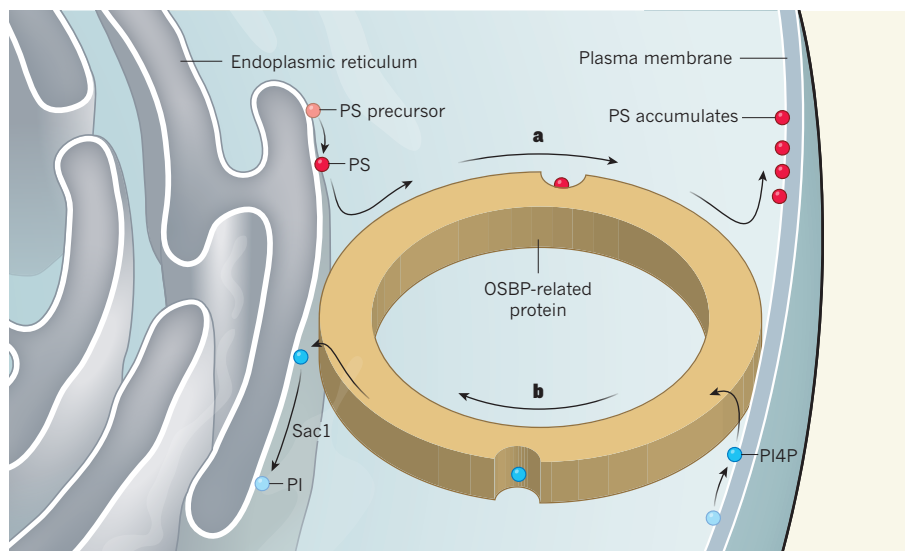
ANANT K. MENON & TIM P. LEVINE

How lipids reach their destinations inside a cell is largely unknown. Most are synthesized in an organelle called the endoplasmic reticulum (ER) and must be delivered to other parts of the cell with speed and specificity. For example, the lipid phosphatidylserine is enriched on the inner surface of the plasma membrane. How does it get there from the ER, and how does it become enriched at the plasma membrane? Two papers in *Science*<sup>1,2</sup> now present a single mechanism that resolves both of these problems.

The membranes that encase the cell and its various organelles each have a unique composition of proteins and lipids. In most instances, proteins are delivered to these membranes by vesicle shuttles that originate in the ER. But many lipids favour non-vesicular modes of transport<sup>3</sup>. These are faster than vesicular trafficking<sup>4</sup> and can produce considerable asymmetries in the lipid compositions of different membranes, leading to lipid enrichment at the plasma membrane compared with the ER.

Because lipids are insoluble in water<sup>4</sup>, during non-vesicular transport they must be shielded from the water-rich cytoplasm by proteins that can hold them in an internal pocket. Such lipid-transfer proteins (LTPs) often act in regions in which the ER comes to within 30 nanometres of another

membrane<sup>4,5</sup>. One class of LTP known to mediate ER-membrane contacts is the family related to oxysterol-binding protein (OSBP), which was originally shown<sup>6</sup> to bind sterols such as oxygenated cholesterol derivatives.



**Figure 1 | Cycling lipid transport.** Two studies<sup>1,2</sup> describe how members of the OSBP protein family move the lipid phosphatidylserine (PS) from its site of synthesis in a cellular organelle called the endoplasmic reticulum (ER) to the inner surface of the plasma membrane, where it becomes enriched. The OSBP-related protein (ORP5 or ORP8 in mammals, Osh6 or Osh7 in yeast) binds PS and another lipid, phosphatidylinositol-4-phosphate (PI4P), in a mutually exclusive fashion. **a**, PS is picked up at the ER and offloaded at the plasma membrane, in exchange for PI4P. **b**, PI4P is delivered to the ER, where the phosphatase enzyme Sac1 hydrolyses it to phosphatidylinositol (PI). The cycle is maintained through the continual resynthesis of PI4P from PI at the plasma membrane.





## 50 Years Ago

There is a growing realization that, in raising productivity ... to the 3.5 per cent national target, efficient utilization of industrial manpower plays a critical part. International comparison with other leading industrial countries implies that British management is wasteful in its use of manpower ... The apparent full employment of the British economy is, in fact, a spurious interpretation of the real employment situation. Present high employment is falsely bolstered by the over-manning of the profitable sectors of the industry, and retention of labour in industries that have ceased to be fully economic ... Without improved data sources or manpower forecast techniques, it is difficult to see how the massive labour movements, between skills, localities of employment or between employers, that the present economic analyses imply to be necessary can be achieved.

From *Nature* 11 September 1965

## 100 Years Ago

The *Engineer* for September 3 has an article on the employment of women as machinists, with special reference to the various shell factories ... Some of these girls ... were found to be capable of a good output on many of the operations after only a week's instruction ... The output on some of the operations exceeded expectation owing to the keenness of the girls, so much so that some of the machines provided have actually been found to be superfluous, and other machines have been found to be capable of more work than had ever been believed to be possible. There is plenty of such labour available in the country, and all the women are moved by the keenest spirit of patriotism. We trust that employers will not hesitate to fill in their blanks from this source.

From *Nature* 9 September 1915

ORD-containing proteins, the sites overlap considerably, meaning that there is room for only one lipid at a time.

*In vitro*, LTPs can facilitate lipid diffusion only down concentration gradients, which *in vivo* would lead to equal concentrations of any given lipid at each membrane. Could the bispecificity of ORD-containing proteins help them to establish asymmetric lipid concentrations? Previous research<sup>8,10</sup> suggests that Osh4 and OSBP mediate the transport of cholesterol from the ER to another organelle, the Golgi, by balancing cholesterol transport in one direction with PI4P transport in the other. PI4P is synthesized in the plasma membrane, Golgi and vesicles called endosomes<sup>11</sup>, and is hydrolysed to phosphatidylinositol by a PI4P phosphatase enzyme called Sac1 that resides in the ER. The lack of PI4P in the ER membrane ensures that Osh4 and OSBP can bind only cholesterol at the ER. On arrival at the Golgi, Osh4 and OSBP offload cholesterol in exchange for more PI4P, allowing the cycle to continue. In this way, even though both cholesterol and PI4P are found at lower levels in the ER than in the Golgi, cholesterol can become enriched in the Golgi<sup>12</sup>.

Could a PI4P countercurrent also drive traffic of phosphatidylserine to the plasma membrane? In one of the new studies, Moser von Filseck *et al.*<sup>2</sup> resolved the structure of the Osh6–PI4P complex (the structure of the Osh6–phosphatidylserine complex is already known<sup>7</sup>), and showed that Osh6 can exchange phosphatidylserine for PI4P between vesicle populations *in vitro*. The ORDs of ORP5 and ORP8 are the closest mammalian counterparts to those of Osh6 and Osh7, and, in the other study, Chung *et al.*<sup>1</sup> showed that the ORD of ORP8 contained phosphatidylserine or PI4P when purified from cells.

Both groups showed that the LTPs were concentrated at contact sites between the ER and the plasma membrane. It is unclear how Osh6 and Osh7 achieve this, but Chung and colleagues found that ORP5 and ORP8 bind to receptors on both membranes. Both studies then showed that these proteins transfer phosphatidylserine from the ER to the plasma membrane, where the proteins pick up PI4P that is subsequently delivered to Sac1. This cycle enriches phosphatidylserine at the plasma membrane (Fig. 1). The hydrolysis of PI4P is a crucial ingredient of this mechanism — the LTPs mediated phosphatidylserine traffic only if they were capable of binding PI4P too, and, in yeast, Sac1 activity was required for phosphatidylserine traffic.

These results will undoubtedly inspire the search for other non-PI4P ligands of proteins related to OSBP, possibly shedding light on the asymmetric transport of many different lipids. Although this is an exciting prospect, the papers raise several questions. First, does Sac1 work exactly as described? Although both papers propose that Sac1 acts after

delivery of PI4P to the ER, other evidence<sup>13</sup> suggests that it acts on PI4P at the plasma membrane.

Second, is a PI4P countercurrent the only way to transport and enrich lipids at the plasma membrane, Golgi and endosomes? Probably not. Cholesterol and phosphatidylserine might be transported by other means<sup>4</sup>, becoming enriched in the plasma membrane through a trapping mechanism. Indeed, the possibility that more than one mechanism might perform this task is supported by the fact that yeast cells lacking Osh6 and Osh7 show no defects<sup>14</sup>. Perhaps the otherwise minor contribution of vesicular transport is increased in these mutants. This could be tested by disabling secretory transport in cells lacking Osh6 and Osh7 or ORP5 and ORP8.

Most data on LTPs have been gathered from *in vitro* experiments, and so the *in vivo* role of these proteins remains enigmatic. Until techniques are developed to determine whether the vast bulk of trafficked lipid molecules are solubilized by an LTP, it remains possible that *in vitro* lipid transfer by LTPs is not matched by the same activity in cells. Instead, LTPs may merely sense lipids, binding them when they are abundant to activate downstream targets that then mediate traffic<sup>13</sup>. Nonetheless, the fact that identically arranged countercurrents pervade not only the ORD-containing family<sup>1,2,8,10,12</sup> but also the Sec14 LTP family<sup>15</sup> is strong evidence that some LTPs can transfer lipids in bulk. ■

Anant K. Menon is in the Department of Biochemistry, Weill Cornell Medical College, New York, New York 10065, USA.

Tim P. Levine is at the UCL Institute of Ophthalmology, University College London, London EC1V 9EL, UK.

e-mails: akm2003@med.cornell.edu; tim.levine@ucl.ac.uk

1. Chung, J. *et al. Science* **349**, 428–432 (2015).
2. Moser von Filseck, J. *et al. Science* **349**, 432–436 (2015).
3. Pagano, R. E. *Curr. Opin. Cell Biol.* **2**, 652–663 (1990).
4. Holthuis, J. C. & Menon, A. K. *Nature* **510**, 48–57 (2014).
5. Prinz, W. A. J. *Cell Biol.* **205**, 759–769 (2014).
6. Olkkonen, V. M. & Li, S. *Prog. Lipid Res.* **52**, 529–538 (2013).
7. Maeda, K. *et al. Nature* **501**, 257–261 (2013).
8. de Saint-Jean, M. *et al. J. Cell Biol.* **195**, 965–978 (2011).
9. Tong, J., Yang, H., Yang, H., Eom, S. H. & Im, Y. J. *Structure* **21**, 1203–1213 (2013).
10. Mesmin, B. *et al. Cell* **155**, 830–843 (2013).
11. Hammond, G. R., Machner, M. P. & Balla, T. J. *Cell Biol.* **205**, 113–126 (2014).
12. Moser von Filseck, J., Vanni, S., Mesmin, B., Antonny, B. & Drin, G. *Nature Commun.* **6**, 6671 (2015).
13. Stefan, C. J. *et al. Cell* **144**, 389–401 (2011).
14. Beh, C. T., Cool, L., Phillips, J. & Rine, J. *Genetics* **157**, 1117–1140 (2001).
15. Kono, N. *et al. Science* **340**, 1106–1110 (2013).

## NEURODEGENERATION

# Amyloid- $\beta$ pathology induced in humans

People who died of the neurodegenerative condition Creutzfeldt–Jakob disease after treatment with cadaver–derived human growth hormone also developed some of the pathological traits of Alzheimer’s disease. [SEE LETTER P.247](#)

MATHIAS JUCKER & LARY C. WALKER

In the 1960s and ’70s, researchers discovered that a rare but deadly human degenerative brain disorder called Creutzfeldt–Jakob disease (CJD) could be transmitted experimentally to animals and, under unusual circumstances, to other humans<sup>1,2</sup>. Since then, some have speculated that other neurodegenerative diseases might also be transmissible<sup>1,3</sup>. On page 247 of this issue, Jaunmuktane *et al.*<sup>4</sup> present evidence indicating that changes in the brain that are characteristic of Alzheimer’s disease have been transmitted between humans. Transmission probably occurred through injections of contaminated, cadaver-derived human growth hormone (c-hGH) that was extracted from pituitary glands collected at autopsy.

Before 1985, an estimated 30,000 people — mostly children with growth deficiency — received injections of c-hGH (refs 2, 5). To obtain sufficient quantities of hormone for treatment, thousands of pituitary glands (a tissue found at the base of the brain) were pooled and homogenized, and c-hGH was then chemically extracted (Fig. 1). After disease incubation times ranging from 5 to more than 40 years, a small percentage of these people (up to 6.3%, according to country<sup>2</sup>) developed CJD. We now know that the CJD-causing contaminant in the pituitary extracts was the prion, a normally produced protein that becomes infectious and toxic by adopting an abnormal shape that similarly corrupts other prion proteins.

Another disease of protein misfolding — and the most prevalent form of dementia — is Alzheimer’s disease<sup>6</sup>. The pathological hallmarks of the disease are insoluble aggregates of amyloid- $\beta$  protein (A $\beta$ ) called plaques, which form between neurons; A $\beta$  build-up in the blood vessels of the brain; and the abnormal deposition of tau protein in nerve cells (known as tauopathy). Several lines of evidence indicate that the misfolding and accumulation of A $\beta$  is an early driver of Alzheimer’s disease, and that this process precedes the onset of dementia by well over a decade<sup>6</sup>. But whether every person with extensive brain A $\beta$  deposition will ultimately develop Alzheimer’s disease is a focus of current research.

It is known that A $\beta$  can aggregate in the brains of animals if their brains are injected with minute amounts of misfolded A $\beta$  proteins known as seeds<sup>7</sup>. This indicates that A $\beta$  deposition can be induced through a prion-like mechanism of corruptive protein templating<sup>7</sup>. By identifying a similar phenomenon in humans, Jaunmuktane and colleagues’ study provides fresh support for this seeding concept in a clinically relevant setting.

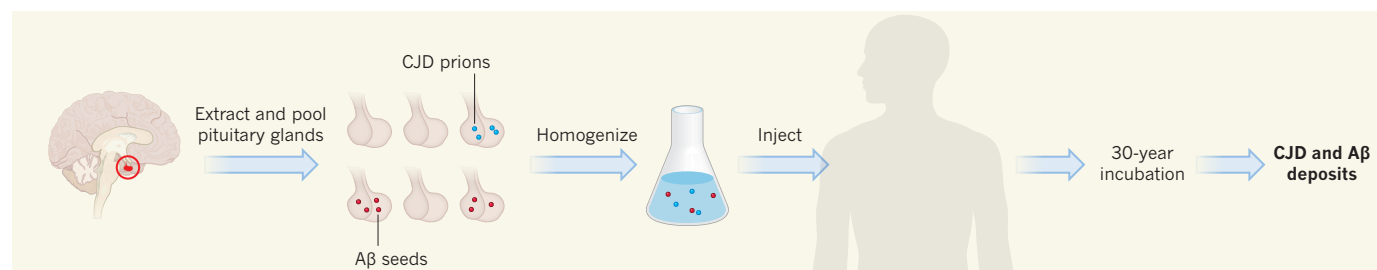
The authors describe the findings of autopsies on 8 people who died of CJD at between 36 and 51 years of age, having been treated with c-hGH approximately 30 years earlier. In addition to the neurodegenerative changes typical of CJD, four of the subjects showed extensive A $\beta$  deposition in the brain and two had sparse A $\beta$  deposits. Such Alzheimer’s-like changes are extremely rare at such a young age, and were

not found in patients up to a decade older who died of prion diseases that were unrelated to c-hGH treatment. The authors also showed that the c-hGH-treated subjects did not have any of the known genetic risk factors for Alzheimer’s disease. Moreover, they confirmed a previous report<sup>8</sup> that A $\beta$  deposits occur in the pituitary glands of people with Alzheimer’s disease, supporting the possibility that aggregates were induced by A $\beta$  seeds in the c-hGH.

Although an observational study such as this cannot prove that the A $\beta$  deposits in the patients’ brains were caused by A $\beta$  seeds, studies in genetically modified mice have established that aggregated A $\beta$  can behave like prions<sup>7,9</sup>. Strikingly, when A $\beta$  seeds were introduced into the abdomens of mice, rather than directly into the brain, A $\beta$  deposition was more prominent in cerebral blood vessels than in A $\beta$  plaques<sup>10</sup>. This finding mirrors the vascular A $\beta$  accumulation observed by Jaunmuktane *et al.*, and reinforces the supposition that the A $\beta$  seeds in the affected people travelled to the brain from elsewhere in the body.

How can future experiments strengthen the case for the prion-like seeding of A $\beta$  in humans and better assess its implications? The original c-hGH extracts, if available, should be assessed for the presence of A $\beta$  seeds using biochemical and animal-transmission experiments. Although age-matched control patients who died of prion disease had a much lower incidence of A $\beta$  deposits than did the patients who died of CJD following c-hGH treatment, there remains a possibility that CJD itself can precipitate Alzheimer’s-like pathology<sup>11</sup>. Understanding the mechanisms by which these different disease processes interact in the brain could help to explain the frequent coexistence of multiple degenerative brain diseases in the elderly<sup>7</sup>.

A $\beta$  seeds are long-lived in the brain, and may be even more resistant to degradation than are prions<sup>12</sup>. Given the build-up of A $\beta$  in the pituitary glands of people with Alzheimer’s disease, and the relatively high prevalence of the disease in the general population, batches of c-hGH are more likely to have been contaminated by A $\beta$  seeds than by prions, which could



**Figure 1 | Contamination of growth-hormone extracts.** Before 1985, people in need of growth-hormone treatment were treated with cadaver-derived human growth hormone (c-hGH). To prepare c-hGH, the pituitary gland at the base of the brain was extracted at autopsy. Of the thousands of glands extracted, a few contained prions from people with the neurodegenerative condition Creutzfeldt–Jakob disease (CJD). Jaunmuktane *et al.*<sup>4</sup> report that some of the

glands probably also contained seeds of amyloid- $\beta$  protein (A $\beta$ ), possibly from people with Alzheimer’s disease. The pooled glands were homogenized and the c-hGH was then extracted and injected into patients. After approximately 30 years, some recipients died of CJD, owing to a build-up of prions. The authors show that some of these people also had A $\beta$  deposits in the brain, suggestive of incipient Alzheimer’s disease.



mean that more recipients received injections containing A $\beta$  seeds. However, it is important to stress that the subjects of this study died of CJD, not of Alzheimer's disease. Whether those with A $\beta$  lesions would eventually have manifested clinical Alzheimer's disease cannot be known with certainty.

Continued surveillance of surviving c-hGH recipients will be essential to determine whether they are at unusually high risk of developing Alzheimer's disease. An earlier study<sup>8</sup> suggests that, as of 2008, c-hGH-treated patients in the United States are not more likely to develop Alzheimer's disease than people in the general population, although an incubation period of 30 years or more is possible. Interestingly, the subjects in the current study lacked tauopathy, an essential feature of Alzheimer's disease<sup>6</sup>. Whether tauopathy would have emerged over a longer incubation period is unknown.

This transmission of A $\beta$  pathology occurred in the uncommon context of long-term c-hGH therapy. So far, there is no indication that Alzheimer's disease can be transmitted between people under ordinary circumstances. Furthermore, the replacement of c-hGH by genetically engineered growth hormone has eliminated the risk that growth-hormone treatment will inadvertently transmit brain disorders between humans. However, it is conceivable that the human transmission of A $\beta$  seeds can occur under other conditions, which must now be carefully defined. Jaunmuktane and colleagues' findings should stimulate new research in this direction, and, more generally, will inspire further investigation into the mechanisms that govern the formation, transmissibility and toxicity of misfolded protein seeds in neurodegenerative diseases. ■

**Mathias Jucker** is at the Hertie Institute for Clinical Brain Research, University of Tübingen, and at the German Center for Neurodegenerative Diseases (DZNE), D-72076 Tübingen, Germany.

**Lary C. Walker** is at the Yerkes National Primate Research Center and Department of Neurology, Emory University, Atlanta, Georgia 30322, USA.

e-mails: mathias.jucker@uni-tuebingen.de; lary.walker@emory.edu

- Gajdusek, D. C. *Science* **197**, 943–960 (1977).
- Brown, P. et al. *Emerg. Infect. Dis.* **18**, 901–907 (2012).
- Prusiner, S. B. *N. Engl. J. Med.* **310**, 661–663 (1984).
- Jaunmuktane, Z. et al. *Nature* **525**, 247–250 (2015).
- Will, R. G. *Br. Med. Bull.* **66**, 255–265 (2003).
- Holtzman, D. M., Morris, J. C. & Goate, A. M. *Sci. Transl. Med.* **3**, 77sr1 (2011).
- Jucker, M. & Walker, L. C. *Nature* **501**, 45–51 (2013).
- Irwin, D. J. et al. *JAMA Neurol.* **70**, 462–468 (2013).
- Prusiner, S. B. *Annu. Rev. Genet.* **47**, 601–623 (2013).
- Eisele, Y. S. et al. *Science* **330**, 980–982 (2010).
- Toussayn, T. et al. *J. Neuropathol. Exp. Neurol.* **74**, 873–888 (2015).
- Ye, L. et al. *Nature Neurosci.* <http://dx.doi.org/10.1038/nn.4117> (2015).

## ATMOSPHERIC SCIENCE

# Sea-spray particles cause freezing in clouds

**Ice clouds in marine regions at high latitudes might form in warmer and drier air than was previously believed because of freezing induced by airborne particles that contain organic materials from ocean surface waters. SEE LETTER P.234**

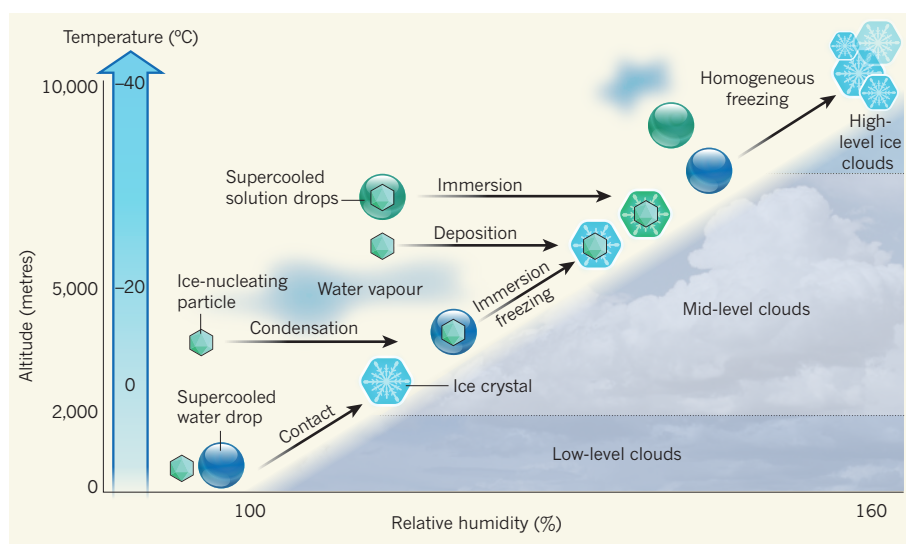
LYNN M. RUSSELL

The oceans cover two-thirds of Earth's surface and are almost entirely, and rather uniformly, composed of water and inorganic salts<sup>1</sup>. The remaining fraction of a per cent of ocean water contains organic material. This has a variable concentration in space and time<sup>2</sup> and is largely uncharacterized, but might be a key component in driving ice formation in the atmosphere. On page 234 of this issue, Wilson *et al.*<sup>3</sup> report that organic material concentrated in the topmost millimetres of the ocean has the essential crystal-forming properties needed to freeze water and form ice clouds in the atmosphere — a process called ice nucleation. The findings might help to refine predictions of future climate.

Ice formation in clouds is central to precipitation processes because it affects whether, when and where rain, snow or ice falls out of clouds. Climate models calculate the timing

and location of ice clouds and the associated precipitation partly on the basis of the particle types and concentrations that are thought to be present in the atmosphere. For example, air temperature must drop to almost  $-40^{\circ}\text{C}$ , and the humidity relative to that at which ice can form at that temperature must be well above 100%, for water to freeze in the atmosphere when no ice-nucleating particles are present<sup>4,5</sup>. But different types of particle can promote freezing when the air is not as cold or as humid as that — by contact with, or immersion in, supercooled water droplets (that is, liquid droplets cooled to below the ideal freezing temperature), by condensation of water onto particles or by direct deposition of ice from water vapour on the particles (Fig. 1).

Wilson *et al.* provide evidence that marine particles could support ice-cloud formation at locations (or at times of the year) where dust is too sparse to freeze ice efficiently. To do this, they sampled surface seawater using a variety of



**Figure 1 | Ice formation in clouds.** The predominant processes for ice formation in the atmosphere depend on temperature (which changes with altitude) and the relative humidity with respect to that at which ice can form. In low-level mixed-phase clouds (composed of water droplets and some ice particles), freezing may occur most effectively when supercooled water droplets come into contact with ice-nucleating particles (INPs). In mid-level mixed-phase and ice clouds, water vapour condenses on INPs, or INPs become immersed in water droplets, after which ice crystals form. Ice crystals can also form when INPs are immersed in supercooled drops of solutions (of salts or of organic compounds, for example), or by direct deposition of ice on the particles. High-level ice clouds include ice that forms ‘homogeneously’ when supercooled droplets freeze or water vapour crystallizes in the absence of INPs. Wilson *et al.*<sup>3</sup> report that particles from the ocean surface can act as INPs. (Figure adapted from refs 4 and 5.)

techniques, used X-ray microscopy to chemically characterize organic material in the water, and observed droplet freezing (both *in situ* and in samples returned to the laboratory).

Bubbles bursting at the ocean surface incorporate some of this surface-ocean organic material into particles that are lofted into the atmosphere, and these particles may have a larger role in forming ice clouds than was previously calculated in climate models. Indeed, Wilson and colleagues show that, when the measured ice-forming abilities of organic materials are represented in a model<sup>6</sup> that calculates the effects of sea-spray particles in global atmospheric simulations, marine particles contribute more to ice nucleation in high-latitude regions, where airborne dust is sparse, than was previously thought. If these results are representative of airborne marine-derived particles around the world, the occurrence of ice clouds in climate simulations could change substantially. The authors' models suggest that the changes will be most evident at high latitudes that have few continents and little desert area, such as the northern Pacific and Atlantic oceans and the Southern Ocean.

Because few measurements of ice-nucleating properties have been taken from ocean surface layers, the model used by the authors necessarily extrapolates the global picture from a limited number of samples in the surface waters of the Arctic and the northern Pacific and Atlantic oceans. To refine things further, it will be necessary to determine the degree to which organic particles from surface waters of, for example, the Southern Ocean differ from marine particles sampled at other latitudes. The simulations could also be improved by characterizing the seasonal and biogeochemical drivers that change the freezing properties of marine organic material and the particles that it forms. Longer-term observations are needed to assess how year-to-year variability in weather and in ocean-nutrient availability affects the formation of organic material that induces freezing.

Wilson and co-workers' findings could also have implications for our understanding of how climate will change in the coming decades. For instance, as global warming occurs, ice clouds might form less frequently in warmer air near the ocean's surface, but stronger surface winds could produce more marine particles to initiate freezing. These two effects may cancel out each other. But if phytoplankton populations decline, then fewer organic ice-freezing particles could be formed, which would exacerbate the reduction in ice-cloud formation.

The authors' work also reveals that marine-derived particles containing organic material were part of the natural mixture of atmospheric particles that made ice freeze in pre-industrial times, but further work is needed to address fundamental questions about marine particles in general: how many of them form,

and what fraction contains ice-freezing organic material? And how do surface winds, ocean ecosystems and the state of the sea change both of these quantities?

Finally, little is known about what controls the size and composition of particles formed as bubbles burst at the ocean surface, but understanding the basic physical processes involved is crucial. Limited measurements and semi-empirical parameterizations provide only a rough basis for climate models to calculate the distribution of such particles in the atmosphere. Satellite observations provide some constraints on the present-day distributions of airborne particles, but without an understanding of the mechanisms of ocean-particle formation, the accuracy and certainty of future

and past contributions from marine particles to changing climate will continue to be limited. ■

**Lynn M. Russell** is at the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0221, USA. e-mail: [lmrussell@ucsd.edu](mailto:lmrussell@ucsd.edu)

1. Holland, H. D. *The Chemistry of the Atmosphere and Oceans* (Wiley, 1978).
2. Hansell, D. A., Carlson, C. A., Repeta, D. J. & Schlitzer, R. *Oceanography* **22**(4), 202–211 (2009).
3. Wilson, T. W. et al. *Nature* **525**, 234–238 (2015).
4. Hoose, C. & Möhler, O. *Atmos. Chem. Phys.* **12**, 9817–9854 (2012).
5. Wendisch, M. & Brenguier, J.-L. (eds) *Airborne Measurements for Environmental Research: Methods and Instruments* (Wiley, 2013).
6. Burrows, S. M., Hoose, C., Pöschl, U. & Lawrence, M. G. *Atmos. Chem. Phys.* **13**, 245–267 (2013).

## ASTROPHYSICS

# Glimpse into a primitive stellar nursery

**The first well-resolved images of local-galaxy stellar nurseries that are poor in elements heavier than helium give the best picture yet of the conditions in which stars may have formed in the early Universe. [SEE LETTER P.218](#)**

ADAM LEROY

Astronomers dub elements heavier than hydrogen and helium 'metals', and these make up only trace amounts of all matter by mass. For example, about 2% of the interstellar matter in the neighbourhood of the Sun comprises metals, the most abundant of which are oxygen and carbon. These elements have a role in catalysing the birth of stars that is far out of proportion to their low abundance. On page 218, Rubio *et al.*<sup>1</sup> present the first well-resolved pictures of metal-deficient stellar nurseries found in a nearby dwarf galaxy, by recording the spectral lines emitted by carbon monoxide (CO). The results open up CO spectroscopic imaging as a diagnostic for exploring the relationship between metal content and star formation for substantially metal-deficient systems.

Stars form out of cold, dense clouds of molecular hydrogen (H<sub>2</sub>). In these clouds, metals act as coolants, helping the gas to reach low temperatures and facilitating its collapse into pre-stellar condensations. Metals also form interstellar dust, which shields stellar nurseries from starlight that would otherwise break molecules apart and heat the gas.

These metals are produced in stellar interiors. When stars die and explode, some of the newly produced metals are mixed into the interstellar gas. Thus, successive generations of stellar birth and death lead to a

gradual enrichment of heavy elements in the interstellar medium. These, in turn, aid the subsequent formation of new stars. Following this logic backwards, early generations of stars probably formed in stellar nurseries that contained few metals compared with the Milky Way or similar present-day galaxies. Thus, to understand the build-up of the first stars and galaxies, astronomers must measure how a dearth of metals (low metallicity) affects the star-formation process.

To study star formation in metal-poor gas, astronomers study the least-massive galaxies in the present-day Universe. These dwarf galaxies are not believed to be truly young, and so they are imperfect analogues of distant primordial systems. But because of a combination of their inefficient star-formation activity and weak gravity (exploding stars can blow heavy elements entirely out of a small galaxy), they are deficient in heavy elements. Therefore, researchers use them as local 'laboratories' to investigate how a lack of metals affects the formation of stars in interstellar gas clouds.

Direct observation of the H<sub>2</sub> that makes up most of the cold, dense gas in these clouds is difficult. This forces astronomers to observe molecular tracers that are mixed with the H<sub>2</sub>, and whose spectral signatures are used to infer the abundance of H<sub>2</sub> indirectly. The workhorse tracer is CO, the second most abundant interstellar molecule<sup>2</sup>. CO survives in the interstellar medium mainly in regions





**Figure 1 | The dwarf galaxy Wolf-Lundmark-Melotte (WLM).** Rubio *et al.*<sup>1</sup> report that the dense clouds containing carbon monoxide in this galaxy (not visible here) have similar physical properties to such clouds in the Milky Way, despite the galaxy's low abundance of elements heavier than helium. The clouds reside inside extended stellar nurseries of hard-to-see cold gas. In this composite image of WLM, young massive stars born in these clouds emit most of the ultraviolet light (blue); optical light (green) is emitted by all stars. The large reservoir of warm, not yet star-forming, interstellar gas emits in radio wavelengths (red).

where there is enough dust to shield it from starlight, and it is studied through its millimetre-wavelength emission, which is detectable by radio telescopes. In a galaxy such as the Milky Way, dust is plentiful and CO is fairly well mixed with H<sub>2</sub>. However, the dust (and the CO itself) is made of heavy elements. Emission from CO therefore tends to be faint in galaxies of low metal content.

There is a long history of hunting for CO in dwarf galaxies, with the goal of understanding metal-poor stellar nurseries. For decades, the Small Magellanic Cloud (a dwarf satellite galaxy of the Milky Way) and a few similar galaxies remained the most metal-poor systems in which stellar nurseries had been detected by their CO emissions. A barrier of about one-fifth of the Milky Way's metallicity emerged<sup>3</sup> as a practical limit to the detection of CO, and direct knowledge of stellar nurseries in galaxies below this limiting value was largely lacking.

Two years ago, researchers from the same group as Rubio *et al.* used the Atacama Pathfinder Experiment telescope in Chile to push past this 'metal barrier'<sup>4</sup>. They observed CO emission from the Wolf-Lundmark-Melotte (WLM) dwarf galaxy, which is part of the same Local Group of galaxies as the Milky Way

(Fig. 1). WLM has a metal content<sup>5</sup> only about one-tenth that of the Milky Way, and about half that of the previous record holder<sup>6</sup> from which CO had been detected. The researchers showed that the CO emission from star-forming regions in WLM was faint compared with that from other tracers of gas and star-formation activity. This implied that CO molecules in WLM traced only the densest, most opaque parts of an extended stellar nursery.

Rubio *et al.* have now used the Atacama Large Millimeter/submillimeter Array (ALMA), the world's most powerful millimetre-wavelength telescope, to take well-resolved pictures of these regions in WLM. The authors' images of CO emission reveal that the nurseries are confined to stunningly small clumps that presumably represent only the densest parts of the star-forming gas clouds (see Fig. 1 of Rubio and co-workers' paper<sup>1</sup>). By contrast, CO emission pervades star-forming regions of the Milky Way, such as the Orion molecular cloud<sup>7,8</sup>. The images enabled the authors to directly measure the CO-emitting clouds' sizes (about 3 parsecs across). They also measured the clouds' kinetic energies, because radio telescopes can track the motion of CO gas by measuring the shift of the frequencies

of its emission lines relative to their rest values (the Doppler effect). On the basis of these measurements, Rubio *et al.* argue that the physical properties (density, pressure and self-gravity) of the CO-containing clouds in the WLM galaxy do resemble those of similarly sized clumps present in metal-rich locales such as the Sun's neighbourhood — even though most of a given star-forming cloud in WLM is invisible in CO emission.

The authors argue that this similarity in physical properties helps to explain why star clusters born in metal-poor galaxies resemble those seen in less-extreme systems. In effect, they propose that the main impact of WLM's lack of metals is to render the bulk of the cloud difficult to see using CO. The lack of dust in WLM means that our best tracer of H<sub>2</sub> is present only deep in the cloud, and the behaviour of most of the H<sub>2</sub> is perhaps not so different from that in other 'normal' galaxies. This agrees, at least qualitatively, with simulations and theoretical predictions for the behaviour of CO and H<sub>2</sub> in metal-poor galaxies<sup>9</sup>. The authors further speculate that the small size of these dust-enshrouded, CO-emitting clumps may explain the relative paucity of highly massive stellar clusters in small, isolated galaxies.

The current study highlights a changing approach to studying star formation in low-metallicity systems. Modern telescopes have begun measuring the energetics, densities and turbulent character of metal-deficient stellar nurseries. This is a substantial advance on simply hunting for faint CO emission from such systems. ALMA is now operating full-time, so we could see exciting progress in this field in the coming years.

However, the fundamental problem of knowing how much H<sub>2</sub> gas is present in metal-poor systems remains daunting, especially given this striking demonstration that CO inhabits only small, dense pockets of gas in the interiors of extended stellar nurseries. High-resolution observations of other gas tracers, including ionized and neutral carbon, and of dust (studied through its infrared emission and the attenuation of starlight that it causes) will be needed to piece together the structure of metal-poor clouds in detail. Rubio *et al.* have investigated these other tracers, but only at poor resolution that is not well matched to the tiny CO-emitting clouds found in WLM.

Finally, it should be noted that only a handful of clouds have been measured in a single system, but the star-formation process can be violent and random on small spatial scales. It will be fascinating to see if these first results are indeed indicative of a broader population of clouds in other low-metallicity galaxies. ■

**Adam Leroy** is in the Department of Astronomy, Ohio State University, Columbus, Ohio 43210, USA.  
e-mail: leroy.42@osu.edu

1. Rubio, M. *et al.* *Nature* **525**, 218–221 (2015).
2. Bolatto, A. D., Wolfire, M. & Leroy, A. K. *Annu. Rev. Astron. Astrophys.* **51**, 207–268 (2013).
3. Taylor, C. L., Kobulnicky, H. A. & Skillman, E. D. *Astron. J.* **116**, 2746 (1998).
4. Elmegreen, B. G. *et al.* *Nature* **495**, 487–489 (2013).

5. Lee, H., Skillman, E. D. & Venn, K. A. *Astrophys. J.* **620**, 223–237 (2005).
6. Israel, F. P., de Graauw, Th., van de Stadt, H. & de Vries, C. P. *Astrophys. J.* **303**, 186 (1986).
7. Wilson, B. A., Dame, T. M., Mashedier, M. R. W. &

- Thaddeus, P. *Astron. Astrophys.* **430**, 523–539 (2005).
8. Buckle, J. V. *et al.* *Mon. Not. R. Astron. Soc.* **401**, 204–222 (2010).
9. Glover, S. C. O. & Clark, P. C. *Mon. Not. R. Astron. Soc.* **426**, 377–388 (2012).
10. Hunter, D. A. *et al.* *Astron. J.* **144**, 134 (2012).

## COMPUTATIONAL BIOLOGY

# How to catch rare cell types

The development of an algorithm called RaceID enables the identification of rare cell types by single-cell RNA sequencing, even when they are part of a complex mixture of similar cells. [SEE LETTER P.251](#)

LU WEN & FUCHOU TANG

How many cell types are there in the human body? Thanks to progress in single-cell sequencing technologies, scientists are now addressing this question in a systematic and non-biased way. On page 251 of this issue, Grün *et al.*<sup>1</sup> take this research forward another step, describing an algorithm called RaceID that can identify rare cell types in a complex mixture of cells.

The transcriptome is the complete collection of RNA molecules present in a cell. Standard approaches to sequencing these assemblages provide an average view of the transcriptome across many cells, and so cannot provide information about differences between individual cells (heterogeneity), or about the characteristics of rare cell types within a heterogeneous population. Such analyses require single-cell transcriptome-sequencing technologies<sup>2–4</sup>, and in the past few years it has become possible to acquire transcriptome data for hundreds and even thousands of single cells<sup>5,6</sup>. Questions have been raised, however, about how reliably information can be mined from these huge data sets, particularly given that they

produce considerable technical noise<sup>7</sup> owing to inaccuracies in the techniques used.

The epithelial cells that line the intestine absorb nutrients and defend the body against microorganisms. The epithelium contains six mature cell types, which are continually renewed by a small population of adult stem cells<sup>8</sup>. This cell layer is one of the best-studied models of self-renewal and differentiation in adult stem cells, and many markers of distinct epithelial cell types have been characterized. This makes it an invaluable system for developing techniques and algorithms for single-cell analysis.

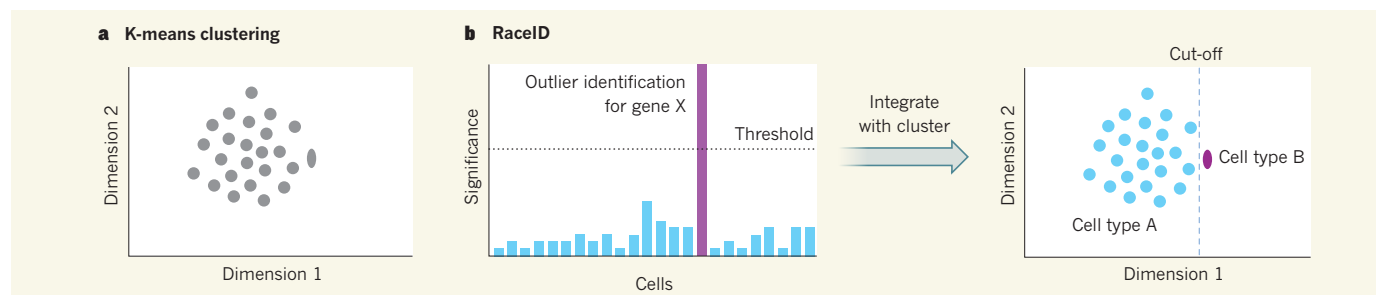
Grün *et al.* used a single-cell transcriptome-sequencing technique to analyse 238 epithelial cells obtained from mouse intestinal organoids — ‘mini guts’ grown *in vitro* from a single stem cell that contain every cell lineage of the intestinal epithelium. Using standard clustering algorithms, the authors distinguished three major cell populations (a rapidly dividing precursor population called transit amplifying cells, absorptive cells called enterocytes and secretory cells). One algorithm, K-means clustering, could distinguish several subgroups within the abundant enterocyte cell population, including

early and late progenitors and mature cells. However, none of the algorithms could distinguish subgroups within the rare secretory-cell lineage, which was represented by only 20 cells in the sample.

The secretory-cell lineage contains at least three cell types, one of which — the hormone-producing enteroendocrine cells — can be further divided into more than ten different subtypes according to the hormones that they secrete<sup>9</sup>. Enteroendocrine cells have key roles in maintaining gut homeostasis, and so an ability to distinguish the different subgroups is desirable. But because of the similarity of their transcriptomes, the subgroups could not be discriminated by standard algorithms in the authors’ initial analyses.

To get around this limitation, Grün *et al.* developed RaceID, a simple and clever algorithm that clearly distinguishes different secretory cell types. The algorithm assumes that a given cell type is likely to strongly express a certain number of cell-type-specific ‘outlier’ genes. Such genes can be identified if care is taken to exclude technical and biological noise (biological noise arises from differences in transcript expression between individual cells of the same type). RaceID identifies outlier cells in each cluster after a K-means clustering step. An outlier cell is defined as expressing a certain number of outlier genes at levels significantly exceeding the modelled noise. In this way, identification of a cell type will not depend on global cell–cell differences, as in standard clustering algorithms, but on only a few genes (Fig. 1).

During single-cell sequencing, each RNA transcript must be amplified many times to provide enough material for accurate sequencing. But the amplification step can introduce technical noise, because small errors in measuring the number of transcripts produced



**Figure 1 | Race to identify rare cells.** **a**, To attempt to differentiate cell types in a mixed population, standard clustering algorithms such as K-means clustering analyse all the RNA transcripts in each cell to determine global differences in gene expression. In this analysis, variations in gene expression between cells are reduced to a two-dimensional space. However, clustering algorithms often fail to identify which cells are of a rare type (oval). **b**, Grün

*et al.*<sup>1</sup> have developed an algorithm, RaceID, that detects outlier genes that are expressed in a cell at a level significantly higher than a given threshold, which is based on the amount of both technical and biological noise. Combining a K-means clustering step with data on which cells express a certain number of outliers above the threshold level identifies cut-off points that enable the identification of rare cell types.



from each gene in a cell are magnified during replication. The authors exclude this noise using a previously reported technique<sup>10</sup> to add a unique molecular 'barcode' to each individual transcript before amplification. This enables the RaceID algorithm to determine whether high levels of gene expression are real or an artefact of amplification. Grün and colleagues demonstrated the efficiency of this strategy using a pool-and-split experiment. They pooled transcripts from 93 cells, split the RNA into 93 equal samples, which created 93 'average' single cells, then amplified and sequenced each sample separately; and no false positive rare cell types were detected.

RaceID identified the gene *Reg4* as being highly expressed specifically in enteroendocrine cells. Grün *et al.* isolated and sequenced a population of 161 *Reg4*-expressing cells. Using RaceID, they identified new enteroendocrine subtypes and validated them *in vivo* at the level of both RNA and protein. This confirmed that RaceID can be used for the identification of rare cell types.

There has been much debate about whether the intestinal stem-cell population is heterogeneous. Can RaceID find subtypes in this population, which is marked by expression of the gene *Lgr5*? Grün and colleagues sequenced transcriptomes from 288 *Lgr5*-expressing cells. RaceID identified these cells as largely homogeneous — the stem-cell population — mixed with a population of rare *Lgr5*-expressing secretory cells. However, as the authors note, it remains possible that the stem-cell population is heterogeneous, but that differences are below a level detectable even by RaceID.

The major limiting factor for RaceID is the accuracy of single-cell sequencing. It is still not possible to measure low-level gene expression accurately in a single cell, and the technical noise for detection of such genes will be too high to identify outliers. The genes for the transcription factors that determine a given cell type are generally not expressed as highly as those encoding hormones, for instance. This might prevent RaceID from discerning potentially functionally important rare cell types in which the differentially expressed genes are likely to mainly encode transcription factors, and may explain the fact that Grün *et al.* were unable to detect stem cells in the initial organoid analysis, because the cells express *Lgr5* at low levels.

The potential for falsely 'identifying' new rare cell types should also be considered. Care must be taken to avoid nucleic-acid cross-contamination or incomplete cell dissociation. It will be necessary to validate putative cell types at the RNA and even protein level.

In terms of sensitivity, accuracy and comprehensiveness, current single-cell sequencing techniques and bioinformatics tools are far from perfect. This is particularly true when it comes to discovering rare cell types. But through the unremitting efforts of Grün *et al.* and others, in the near future we

may be able to chart a complete cell-lineage map of the human body. ■

**Lu Wen and Fuchou Tang** are at the *Biodynamic Optical Imaging Center, School of Life Sciences, Peking University, Beijing 100871, China.*

*e-mails: wenlu.wl@gmail.com; tangfuchou@pku.edu.cn*

1. Grün, D. *et al.* *Nature* **525**, 251–255 (2015).

## CANCER

# A moving target

**An *in silico*, three-dimensional model of tumour evolution suggests that cell motility is a key factor in the initial growth of a tumour mass. The model also reveals the dynamics of mutation spread. SEE LETTER P.261**

**NATALIA L. KOMAROVA**

Evolutionary thinking is becoming an indispensable tool to understand cancer, and even to propose directions in the search for treatment strategies. In this issue, Waclaw *et al.*<sup>1</sup> (page 261) use mathematical modelling based on evolutionary principles to provide an explanation for the observed architecture of tumours, and to argue that cell migration might be the key to tumour shape, spread and drug resistance. This study opens up the possibility of treatments that target genes related to cell motility and adhesion, rather than the conventional targets of genes governing cell division, death and differentiation.

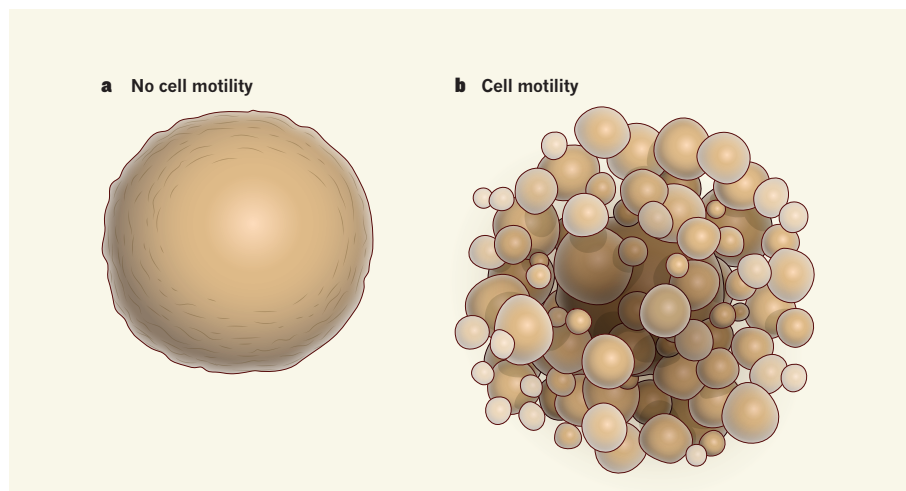
Cancer is an unwanted evolutionary process whereby cells, driven by random mutations,

2. Tang F., Lao, K. & Surani, M. A. *Nature Methods* **8**, S6–S11 (2011).
3. Eberwine, J., Sul, J. Y., Bartfai, T. & Kim, J. *Nature Methods* **11**, 25–27 (2014).
4. Treutlein, B. *et al.* *Nature* **509**, 371–375 (2014).
5. Klein, A. M. *et al.* *Cell* **161**, 1187–1201 (2015).
6. Macosko, E. Z. *et al.* *Cell* **161**, 1202–1214 (2015).
7. Stegle, O., Teichmann, S. A. & Marioni, J. C. *Nature Rev. Genet.* **16**, 133–145 (2015).
8. Clevers, H. *Cell* **154**, 274–284 (2013).
9. May, C. L. & Kaestner, K. H. *Mol. Cell. Endocrinol.* **323**, 70–75 (2010).
10. Jaitin, D. A. *et al.* *Science* **343**, 776–779 (2014).

This article was published online on 19 August 2015.

escape the orchestrated behaviour of a functioning tissue and enter a phase of abnormal growth and, later, metastasis (tumour spread). We still lack understanding of many aspects of this complex process, and researchers in different fields are collaborating to solve this ultimate riddle. Evolutionary biologists approach the study of tumours in a similar manner to the study of viruses, bacteria or animals: they seek the mutations that give rise to the ever-changing variety of tumour cells, and they look at the forces of natural selection that allow certain mutants to proliferate, replace their competitors and give rise to new waves of evolutionary change.

Waclaw *et al.* combined methods from evolutionary biology and ecology with current knowledge of the molecular biology of cancer to design a versatile mathematical model



**Figure 1 | Motility helps to explain tumour dynamics.** Waclaw *et al.*<sup>1</sup> present an *in silico* model of 3D tumour architecture over the course of tumour evolution. **a**, When cells in the model are unable to move, a slow-growing tumour of largely spherical shape is predicted. **b**, By contrast, motile cells lead to a faster-growing tumour with a conglomerate structure more similar to that seen in clinical examples.

in which tumour-cell populations undergo evolutionary change, guided by realistic parameters. The authors used this model to study the growth laws of 3D *in silico* tumours, focusing on the tumours' geometry and cellular composition. If the tumour cells are relatively immobile, then, as they proliferate and form a malignant mass, they crowd each other out and thus slow their own replication. Soon the tumour can grow only at its surface, which results in a relatively slow expansion (the mass grows as a cubic power of time).

But this slow growth cannot account for the relatively fast tumour expansion observed in many clinical studies. The resolution of this apparent paradox lies in cellular motility. By giving each cell in the model the ability to migrate, the researchers observed a much faster, exponential, growth, which also yielded a different, more realistic, tumour shape (Fig. 1). This result is consistent with the earlier proposition<sup>2</sup> that migratory potential is a component of a cell's evolutionary fitness, in much the same way as is its replicative potential. However, it was previously thought that cell migration was mostly involved in the invasion of tissues by tumours or in metastasis. The direct, pivotal role of cell motility in tumour growth was under-appreciated and can now be considered a valid treatment target.

Another focus of the authors' study was tumour composition. In particular, they asked how quickly a particular mutation can propagate in a mass of cancer cells, thus changing the tumour's properties. Evolutionary processes and their outcomes are largely shaped by the environment in which they take place. For example, evolution in a well-mixed, homogeneous medium takes place at a different pace from evolution in an environment in which interactions are restricted by geometric constraints. And in the latter case, dimensionality is key. For example, it has been shown that inactivation of a tumour-suppressor gene (a two-hit evolutionary process in which the cells must first become less fit before becoming more fit) happens faster in 1D (a row of cells)<sup>3,4</sup> than in 2D (a layer), and this is in turn faster than in a fully mixed system with no spatial constraints<sup>4–7</sup>. By contrast, in two-step processes in which the intermediate mutant confers a slight selective advantage, the relationship is the opposite, and a non-spatial, fully mixed environment promotes the fastest pace of evolution<sup>5</sup>. These phenomena seem less surprising if one notes how reminiscent they are of other fundamental laws of nature in which space dimensionality changes how things work, such as the different fundamental solutions of Poisson's equations in 1D and 2D.

Waclaw *et al.* then set out to understand why, given the high overall degree of tumour heterogeneity, some mutations are so prevalent among the cells of a given tumour. In the context of tumour progression, two broad classes

of mutation have been identified<sup>8</sup>. Cells with driver mutations are characterized by having a growth advantage over other cells, and such mutations are thought to be responsible for cancer initiation and progression. Passenger mutations are genomic changes that do not really alter the cells' growth properties, and do not have a causal role in cancer origin or progression. Waclaw and colleagues show that, in the presence of even a small amount of selective advantage (that is, a driver mutation), the affected cells sweep rapidly through a 3D cell population. This explains the observed composition of large tumours, in which almost every cell contains the same driver mutations, and heterogeneity resulting from passenger mutations accumulates later, during tumour progression.

This idea is crucial in the context of cancer therapy. A mutation that confers resistance to a drug is usually a passenger mutation before treatment; such mutations are 'hiding' inside any tumour and are generated simply by chance as a result of the constant background mutation rate. But resistant mutants immediately gain a selective advantage once treatment is applied. Waclaw and colleagues' paper illustrates how rapidly resistant cells can accumulate, leading to regrowth and treatment failure. This happens even faster in the presence of mutations that increase cellular motility.

How far are we from being able to use evolution to our advantage? Understanding evolution's intricate ways brings us a step closer to being able to reverse malignant processes, and to channel the dynamics in the direction we want. And can we use the genes responsible for cell motility or cell adhesion as targets in future cancer treatments? Waclaw and colleagues' theoretical study suggests that this is a possibility, and it is to be hoped that others will take up the challenge. ■

**Natalia L. Komarova** is in the Departments of Mathematics and Ecology and Evolutionary Biology, University of California, Irvine, Irvine, California 92697, USA.  
e-mail: komarova@uci.edu

1. Waclaw, B. *et al.* *Nature* **525**, 261–264 (2015).
2. Thalhauser, C. J., Lowengrub, J. S., Stupack, D. & Komarova, N. L. *Biol. Direct* **5**, 21 (2010).
3. Michor, F., Iwasa, Y., Rajagopalan, H., Lengauer, C. & Nowak, M. A. *Cell Cycle* **3**, 356–360 (2004).
4. Komarova, N. L. *Bull. Math. Biol.* **68**, 1573–1599 (2006).
5. Komarova, N. L., Shahriyari, L. & Wodarz, D. J. R. *Soc. Interface* **11**, 20140014 (2014).
6. Durrett, R. & Moseley, S. *Ann. Appl. Probab.* **25**, 104–115 (2015).
7. Komarova, N. L. *Proc. Natl Acad. Sci. USA* **111**, 10789–10795 (2014).
8. Haber, D. A. & Settleman, J. *Nature* **446**, 145–146 (2007).

This article was published online on 26 August 2015.

## CANCER

# Mutant p53 and chromatin regulation

**The finding that genes encoding enzymes that modify histone proteins are among the targets of certain mutant forms of the p53 protein sheds light on how these mutations cause cancer beyond p53 inactivation. [SEE ARTICLE P.206](#)**

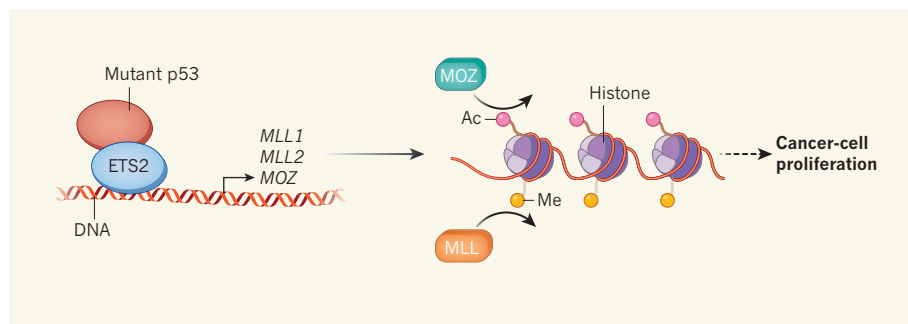
CAROL PRIVES & SCOTT W. LOWE

**M**utations in the *TP53* tumour-suppressor gene are common in human tumours. Although these mutations invariably inactivate the normal activity of p53 (ref. 1), which is the transcription factor encoded by *TP53*, some mutations also endow p53 with 'gain-of-function' activities that promote cancer<sup>2</sup>. Whether diverse p53 mutants produce similar gain-of-function activities and how they do so remains a puzzle, but finding the answer might enable the design of strategies for treating many cancers. On page 206 of this issue, Zhu *et al.*<sup>3</sup> provide a possible explanation: they show that gain-of-function mutant p53 proteins induce the production of enzymes that modify the histone proteins around which DNA is packaged

as chromatin, thus altering gene expression.

Experimentally altering the expression of gain-of-function mutant p53 affects the expression of myriad genes, enhancing the invasiveness and proliferation of tumour cells *in vitro*<sup>4</sup>. Moreover, mice harbouring key gain-of-function mutations in *TP53* develop tumours that differ from those lacking p53 (ref. 5). Lowering the level of gain-of-function p53 has antiproliferative effects *in vitro* and can reduce metastasis or trigger tumour regression *in vivo*<sup>2,5,6</sup>. A better understanding of these mutants is therefore desirable. Zhu *et al.* found that, in cultured human-cancer cell lines, gain-of-function mutant forms of p53 bind to different regions of DNA from the normal protein. In particular, the mutant proteins bind to the genes *MLL1* and *MLL2*. Gain-of-function p53 seems to be recruited to





**Figure 1 | Gaining on p53.** ‘Gain-of-function’ mutations in the tumour-suppressor gene *TP53* enable the transcription factor that it encodes to bind to abnormal targets, leading to cancer. Zhu *et al.*<sup>3</sup> report that gain-of-function p53 binds to the transcription factor ETS2 and activates the genes *MLL1*, *MLL2* and *MOZ*. *MLL1* and *MLL2* encode MLL enzymes that add methyl groups (Me) to the histone proteins around which genes are packaged as chromatin, and MOZ adds acetyl groups (Ac) to these histones. Both modifications increase local gene expression, leading to an increase in the proliferation of cancer cells through as-yet-unknown mechanisms.

these genes in part through binding to ETS2 — a transcription factor that is known<sup>7</sup> to target gain-of-function p53 to different genes from those activated by normal p53.

*MLL1* and *MLL2* are members of the SET family of histone methyltransferase enzymes<sup>8</sup>. These act as parts of large complexes<sup>9</sup> to modulate gene expression by attaching methyl groups to a lysine amino-acid residue (K4) of the histone H3 protein. Such H3K4 methylation allows increased transcription of the gene packaged around the histones. The authors found that gain-of-function p53 also activates expression of the gene *MOZ*, which encodes an enzyme that adds an acetyl group to K9 of H3, again allowing increased gene expression.

In agreement with the idea that gain-of-function p53 affects histone modification, reducing levels of the mutant p53 decreased H3K9 acetylation. However, p53 reduction had only a small effect on H3K4 methylation. Perhaps, as Zhu and colleagues suggest, this is because other members of the SET family have similar but p53-independent roles to *MLL1* and *MLL2*. The authors demonstrated that gain-of-function p53 activates *MLL1*, *MLL2* and *MOZ*, and showed that this activation is partly responsible for the ability of the mutant p53 to enhance cell proliferation *in vitro* (Fig. 1).

Finally, mining human-cancer databases provided support for Zhu and colleagues’ data, indicating that expression of *MLL1*, *MLL2* and *MOZ* is significantly upregulated in human tumours with select p53 gain-of-function mutants compared with tumours lacking p53 or those without mutant p53. This correlation is not obvious across breast cancers<sup>10</sup>, the tissue of origin for many cell lines studied in this work — although Zhu *et al.* clearly show that such a correlation exists in the cell lines that they used. It is likely that other variables affect MLL expression in some tumour types. As a result, it will be important to investigate the contextual factors that determine

whether or when gain-of-function p53 can trigger changes in *MLL* and *MOZ* expression, and to analyse the mechanisms underlying these events.

Gain-of-function p53 was also recently shown<sup>11</sup> to act with the SWI/SNF chromatin-remodelling complex to upregulate many genes that can themselves mediate the cancer-causing activities of gain-of-function p53. This finding, taken together with Zhu and colleagues’ demonstration that this p53 is linked to chromatin and, by extension, to the transcriptome (the complete gene-expression profile of the cell), could explain why so many genes are affected by the presence of gain-of-function p53. But precisely how p53 proteins with diverse mutations acquire similar capabilities remains to be discovered.

One possibility is that p53 mutants adopt a different structure from normal p53 that enables their interaction with ETS2. However, there is no obvious explanation for the evolution of such an interaction, and this model is at odds with the observation<sup>12</sup> that some gain-of-function p53 proteins have similar structures to normal p53. Alternatively, the ability of normal p53 to bind to thousands of sites in the human genome might prevent it from associating with the factors with which gain-of-function p53 interacts. Or perhaps the expression of one or more target genes somehow actively prevents the normal protein from engaging in the interactions that are characteristic of the mutant p53.

The authors’ link between gain-of-function p53 and *MLL1* and *MLL2* is intriguing, given that members of the *MLL* family are frequently mutated in human cancers<sup>13</sup>. For example, chromosomal translocations involving *MLL1* can drive leukaemia, and *MLL2* mutations are common in several carcinomas. But *MLL1* translocations eliminate the gene’s histone methyltransferase domain, and *MLL2* mutations seem to be inactivating. The explanation for this apparent discrepancy with Zhu and colleagues’ findings is unclear,

but probably reflects context-dependent differences in enzyme function.

Could targeting *MLL1*, *MLL2* or *MOZ* be a strategy for treating tumours involving gain-of-function p53? Zhu *et al.* showed that two inhibitors of MLL-complex formation block proliferation in cells expressing mutant p53, but do not affect those lacking p53. Eliminating gain-of-function p53 or interfering with its mechanism of action can have anticancer effects *in vitro* and in mice<sup>6,14,15</sup>. Moreover, there is much enthusiasm for cancer treatments that affect chromatin modification, and compounds that target some chromatin-modifying activities have been approved for use in the clinic or are currently in clinical trials. But more work is required, because the specificity of the MLL inhibitors is not entirely established. Furthermore, *MLL* genes are active during embryonic development, and their inhibition can cause embryonic death, independent of *TP53* mutations<sup>14</sup>. These observations, together with the previously mentioned fact that mutations disrupting MLL function are common in tumours, raise concerns that MLL inhibitors might be toxic, or might even promote tumours.

Nonetheless, given the frequency with which *TP53* is mutated in cancer, continued efforts to modulate the effects of mutant p53 are clearly warranted. With the key targets of MLL and MOZ in hand, specific therapies might become possible. Thus, Zhu and colleagues’ study, and those of others, might point to treatments for tumours that harbour *TP53* mutations. ■

Carol Prives is in the Department of Biological Sciences, Columbia University, New York, New York 10027, USA.

Scott W. Lowe is at the Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.  
e-mails: clp3@columbia.edu;  
lowes@mskcc.org

1. Vogelstein, B. & Kinzler, K. W. *Cell* **70**, 523–526 (1992).
2. Muller, P. A. J. & Vousden, K. H. *Nature Cell Biol.* **15**, 2–8 (2013).
3. Zhu, J. *et al.* *Nature* **525**, 206–211 (2015).
4. Freed-Pastor, W. A. & Prives, C. *Genes Dev.* **26**, 1268–1286 (2012).
5. Garcia, P. B. & Attardi, L. D. *Semin. Cell Dev. Biol.* **27**, 74–85 (2014).
6. Alexandrova, E. M. *et al.* *Nature* **523**, 352–356 (2015).
7. Do, P. M. *et al.* *Genes Dev.* **26**, 830–845 (2012).
8. Zhang, P., Bergamin, E. & Couture, J.-F. *Biopolymers* **99**, 136–145 (2013).
9. Shilatifard, A. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
10. www.cBioPortal.org
11. Pfister, N. T. *et al.* *Genes Dev.* **29**, 1298–1315 (2015).
12. Joerger, A. C., Ang, H. C. & Fersht, A. R. *Proc. Natl Acad. Sci. USA* **103**, 15056–15061 (2006).
13. Ford, D. J. & Dingwall, A. K. *Cancer Genet.* **208**, 178–191 (2015).
14. Freed-Pastor, W. A. *et al.* *Cell* **148**, 244–258 (2012).
15. Weissmueller, S. *et al.* *Cell* **157**, 382–394 (2014).

This article was published online on 2 September 2015.

# Mapping tree density at a global scale

T. W. Crowther<sup>1</sup>, H. B. Glick<sup>1</sup>, K. R. Covey<sup>1</sup>, C. Bettigole<sup>1</sup>, D. S. Maynard<sup>1</sup>, S. M. Thomas<sup>2</sup>, J. R. Smith<sup>1</sup>, G. Hintler<sup>1</sup>, M. C. Duguid<sup>1</sup>, G. Amatulli<sup>3</sup>, M.-N. Tuanmu<sup>3</sup>, W. Jetz<sup>1,3,4</sup>, C. Salas<sup>5</sup>, C. Stam<sup>6</sup>, D. Piotto<sup>7</sup>, R. Tavani<sup>8</sup>, S. Green<sup>9,10</sup>, G. Bruce<sup>9</sup>, S. J. Williams<sup>11</sup>, S. K. Wiser<sup>12</sup>, M. O. Huber<sup>13</sup>, G. M. Hengeveld<sup>14</sup>, G.-J. Nabuurs<sup>14</sup>, E. Tikhonova<sup>15</sup>, P. Borchardt<sup>16</sup>, C.-F. Li<sup>17</sup>, L. W. Powrie<sup>18</sup>, M. Fischer<sup>19,20</sup>, A. Hemp<sup>21</sup>, J. Homeier<sup>22</sup>, P. Cho<sup>23</sup>, A. C. Vibrans<sup>24</sup>, P. M. Umunay<sup>1</sup>, S. L. Piao<sup>25</sup>, C. W. Rowe<sup>1</sup>, M. S. Ashton<sup>1</sup>, P. R. Crane<sup>1</sup> & M. A. Bradford<sup>1</sup>

**The global extent and distribution of forest trees is central to our understanding of the terrestrial biosphere. We provide the first spatially continuous map of forest tree density at a global scale. This map reveals that the global number of trees is approximately 3.04 trillion, an order of magnitude higher than the previous estimate. Of these trees, approximately 1.39 trillion exist in tropical and subtropical forests, with 0.74 trillion in boreal regions and 0.61 trillion in temperate regions. Biome-level trends in tree density demonstrate the importance of climate and topography in controlling local tree densities at finer scales, as well as the overwhelming effect of humans across most of the world. Based on our projected tree densities, we estimate that over 15 billion trees are cut down each year, and the global number of trees has fallen by approximately 46% since the start of human civilization.**

Forest ecosystems harbour a large proportion of global biodiversity, contribute extensively to biogeochemical cycles, and provide countless ecosystem services, including water quality control, timber stocks and carbon sequestration<sup>1–4</sup>. Our current understanding of the global forest extent has been generated using remote sensing approaches that provide spatially explicit values relating to forest area and canopy cover<sup>3,5,6</sup>. Used in a wide variety of global models, these maps have enhanced our understanding of the Earth system<sup>3,5,6</sup>, but they do not currently address population numbers, densities or timber stocks. These variables are valuable for the modelling of broad-scale biological and biogeochemical processes<sup>7–9</sup> because tree density is a prominent component of ecosystem structure, governing elemental processing and retention rates<sup>7,9,10</sup>, as well as competitive dynamics and habitat suitability for many plant and animal species<sup>11–13</sup>.

The number of trees in a given area can also be a meaningful metric to guide forest management practices and inform decision-making in public and non-governmental sectors<sup>14,15</sup>. For example, international afforestation efforts such as the ‘Billion Trees Campaign’, and city-wide projects including the numerous ‘Million Tree’ initiatives around the world have motivated civil society and political leaders to promote environmental stewardship and sustainable land management by planting large numbers of trees<sup>14,16,17</sup>. Establishing targets and evaluating the proportional contribution of such projects requires a sound baseline understanding of current and potential tree population numbers at regional and global scales<sup>16,17</sup>.

The current estimate of global tree number is approximately 400.25 billion<sup>18</sup>. Generated using satellite imagery and scaled based on global forest area, this estimate engaged policy makers and environmental practitioners worldwide by suggesting that the ratio of trees-to-people is 61:1. This has, however, been thrown into doubt by a recent broad-scale inventory that used 1,170 ground-truthed measurements of tree density to estimate that there are 390 billion trees in the Amazon basin alone<sup>19</sup>.

## Mapping tree density

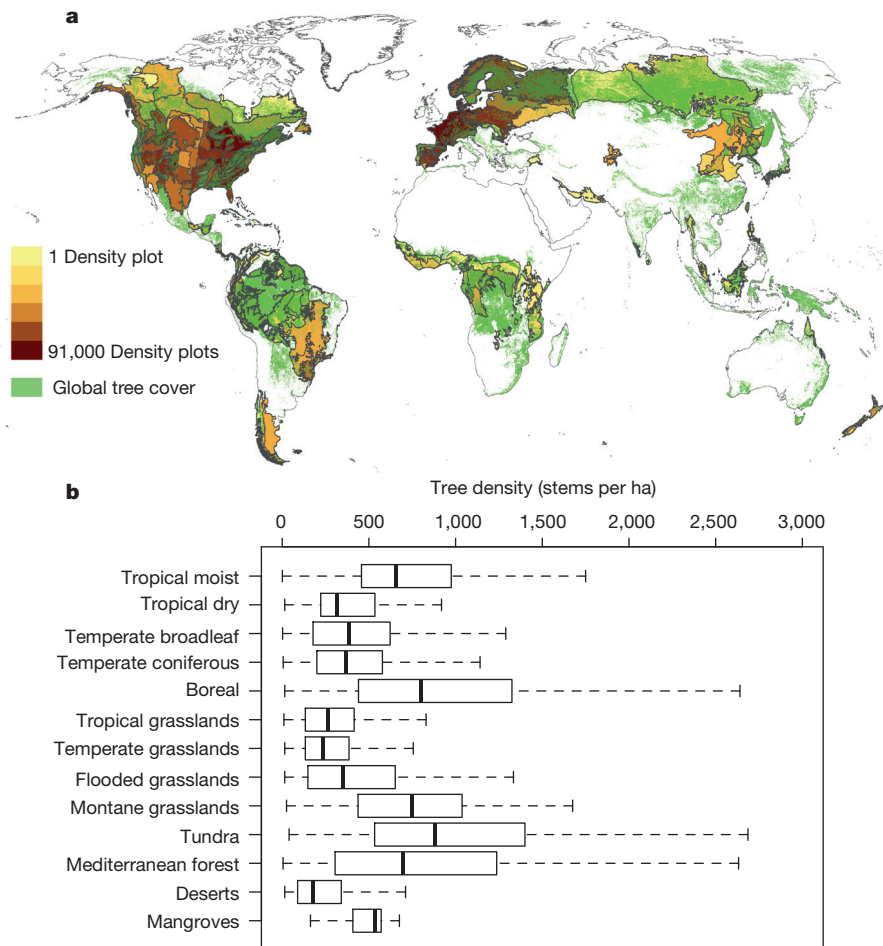
Here, we use 429,775 ground-sourced measurements of tree density from every continent on Earth except Antarctica to generate a global map of forest trees. Forested areas are found in most of Earth’s biomes, even those as counterintuitive as desert, tundra, and grassland (Fig. 1a, b). We generated predictive regression models for the forested areas in each of the 14 biomes as defined by The Nature Conservancy (<http://www.nature.org>). These models link tree density to spatially explicit remote sensing and geographic information systems (GIS) layers of climate, topography, vegetation characteristics and anthropogenic land use (see Extended Data Table 1). Following almost all of the collected data sources, we define a tree as a plant with woody stems larger than 10 cm diameter at breast height (DBH)<sup>19</sup>.

Incorporating plot-level measurements from more than 50 countries, the measured tree density values were inherently variable within and among biomes (Figs 1 and 2). However, the large number of tree density measurements ensured that the confidence in our mean (and total) estimates is high (Fig. 3). Furthermore, the scale of these data

<sup>1</sup>Yale School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511, USA. <sup>2</sup>Department of Environmental Sciences, University of Helsinki, Helsinki 00014, Finland.

<sup>3</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06511, USA. <sup>4</sup>Department of Life Sciences, Silwood Park, Imperial College, London SL5 7PY, UK. <sup>5</sup>Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco 4811230, Chile. <sup>6</sup>RedCastle Resources, Salt Lake City, Utah 84103, USA. <sup>7</sup>Universidade Federal do Sul da Bahia, Ferradas, Itabuna 45613-204, Brazil. <sup>8</sup>Forestry Department, Food and Agriculture Organization of the United Nations, Rome 00153, Italy. <sup>9</sup>Operation Wallacea, Spilsby, Lincolnshire PE23 4EX, UK. <sup>10</sup>Durrell Institute of Conservation and Ecology (DICE), School of Anthropology and Conservation (SAC), University of Kent, Canterbury ME4 4AG, UK. <sup>11</sup>Molecular Imaging Research Center MIRcen/CEA, CNRS URA 2210, 91401 Orsay Cedex, France. <sup>12</sup>Landcare Research, Lincoln 7640, New Zealand. <sup>13</sup>WSL, Swiss Federal Institute for Forest, Snow and Landscape Research, 8903 Birmensdorf, Switzerland. <sup>14</sup>Environmental Science Group, Wageningen University & Research Centre, 6708 PB, The Netherlands. <sup>15</sup>Center for Forest Ecology and Productivity RAS, Moscow 117997, Russia. <sup>16</sup>CEN Center for Earth System Research and Sustainability, Institute of Geography, University of Hamburg, Hamburg 20146, Germany. <sup>17</sup>Department of Botany and Zoology, Masaryk University, Brno 61137, Czech Republic. <sup>18</sup>South African National Biodiversity Institute, Kirstenbosch Research Centre, Claremont 7735, South Africa. <sup>19</sup>Institute of Plant Sciences, Botanical Garden, and Oeschger Centre for Climate Change Research, University of Bern, 3013 Bern, Switzerland. <sup>20</sup>Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre (BKC-F), 60325 Frankfurt, Germany. <sup>21</sup>Department of Plant Systematics, University of Bayreuth, 95447 Bayreuth, Germany. <sup>22</sup>Albrecht von Haller Institute of Plant Sciences, Georg August University of Göttingen, 37073 Göttingen, Germany. <sup>23</sup>Tropical Ecology Research Group, Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. <sup>24</sup>Universidade Regional de Blumenau, Departamento de Engenharia Florestal, Blumenau/Santa Catarina 89030-000, Brazil. <sup>25</sup>Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China.



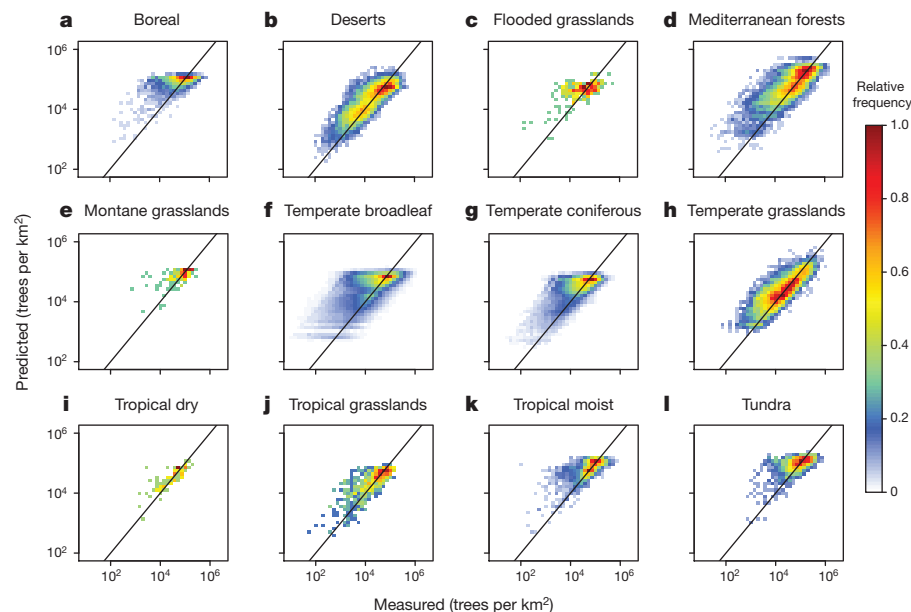


**Figure 1 | Map of data points and raw biome-level forest density data.** **a**, Image highlighting the ecoregions (shapefiles provided by The Nature Conservancy (<http://www.nature.org>)) from which the 429,775 ground-sourced measurements of tree density were collected. Shading indicates the

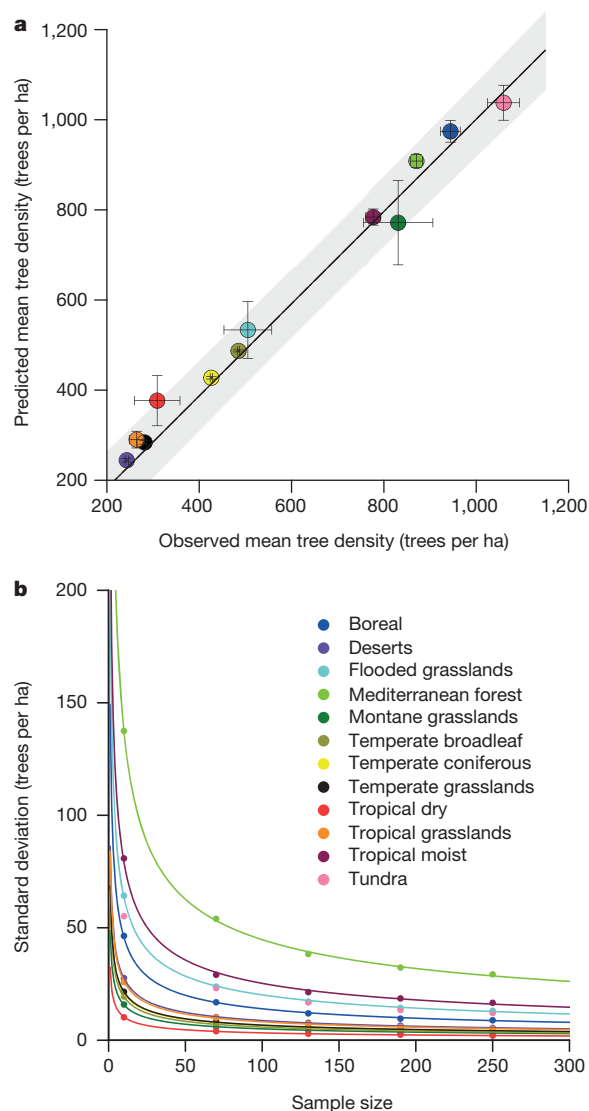
total number of plot measurements collected in each ecoregion. A global forest map was overlaid in green to highlight that collected data span the majority of forest ecosystems on a global scale. **b**, The median and interquartile range of tree density values collected in the forested areas of each biome.

ensures that our modelled estimates are unlikely to be influenced significantly by recent forest loss, reforestation or natural forest regeneration, which are responsible for a net global change of <1% of the

global forest area each year<sup>3</sup>. Biome-level validation estimates indicate that our models have high precision when predicting the mean tree densities of omitted validation plots (Fig. 3a). Although the accuracy



**Figure 2 | Heat plots showing the relationships between predicted and measured tree density data.** **a–l**, Predictions were generated using generalized linear models ( $n = 429,775$ ). Diagonal lines indicate 1:1 lines (perfect correspondence) between predicted and observed points, scaled to the kilometre level. Colours indicate the proportion of data points from that biome that fall within each pixel. Biomes with a greater number of plot measurements have greater variability but higher confidence in the mean estimates, highlighting the trade-off between broad-scale precision and fine-scale accuracy. Axes are log-transformed to account for exceptionally high variability in tree density.



**Figure 3 | Validation plots for biome-level predictions.** **a**, Biome-level regression models predict the mean values of the omitted validation plot measurements in 12 biomes. Overall, the models underestimated mean tree density by  $\sim 3\%$  (slope = 0.97) but this difference was not statistically significant ( $P = 0.51$ ). Bars show  $\pm$  one standard deviation for the predicted mean and the grey area represents the 95% confidence interval for the mean. The values plotted here represent mean densities for the plot measurements (that is, for forested ecosystems), rather than those predicted for each entire biome. **b**, The standard deviation of the predicted mean values as a function of sample size. As sample size increases, the variability of the predicted mean tree density reaches a threshold, beyond which an increase in sample size results in a minimal increase in precision. Standard deviations were calculated using a bootstrapping approach (see Methods), and smooth curves were modelled using standard linear regression with a log–log transformation.

of our models is limited at the level of an individual hectare, the precision of the mean density estimates is high ( $\pm 40$  trees  $\text{ha}^{-1}$ ) beyond a threshold of  $\sim 200$  plots (Fig. 3b).

### Global-level and biome-level patterns

Together, the biome-level models provide the first spatially continuous map of global tree densities at a  $1\text{-km}^2$  (30 arc-seconds) resolution (Fig. 4a). Based on this map, we estimate that the global number of trees is approximately 3.04 trillion ( $\pm 0.096$  trillion, 95% confidence intervals (CI)). An order of magnitude higher than the previous global estimate<sup>18</sup>, the scale of our projection is consistent with recent large-scale inventories

in Europe, North America and the Amazon basin<sup>19</sup> (Fig. 4d). With a human population of 7.2 billion, our estimate of global tree density revises the ratio of trees per person from 61:1 to 422:1.

At the biome-level, the highest tree densities exist in forested regions of the Boreal and Tundra zones (Fig. 1b). In these northern latitudes, limited temperature and moisture lead to the establishment of stress-tolerant coniferous tree species that can reach the highest densities on Earth (Fig. 1). However, the tropical regions contain a greater proportion of the world's forested land. A total of 42.8% of the planet's trees exist in tropical and subtropical regions, with another 24.2% and 21.8% in boreal and temperate biomes, respectively (Fig. 4a).

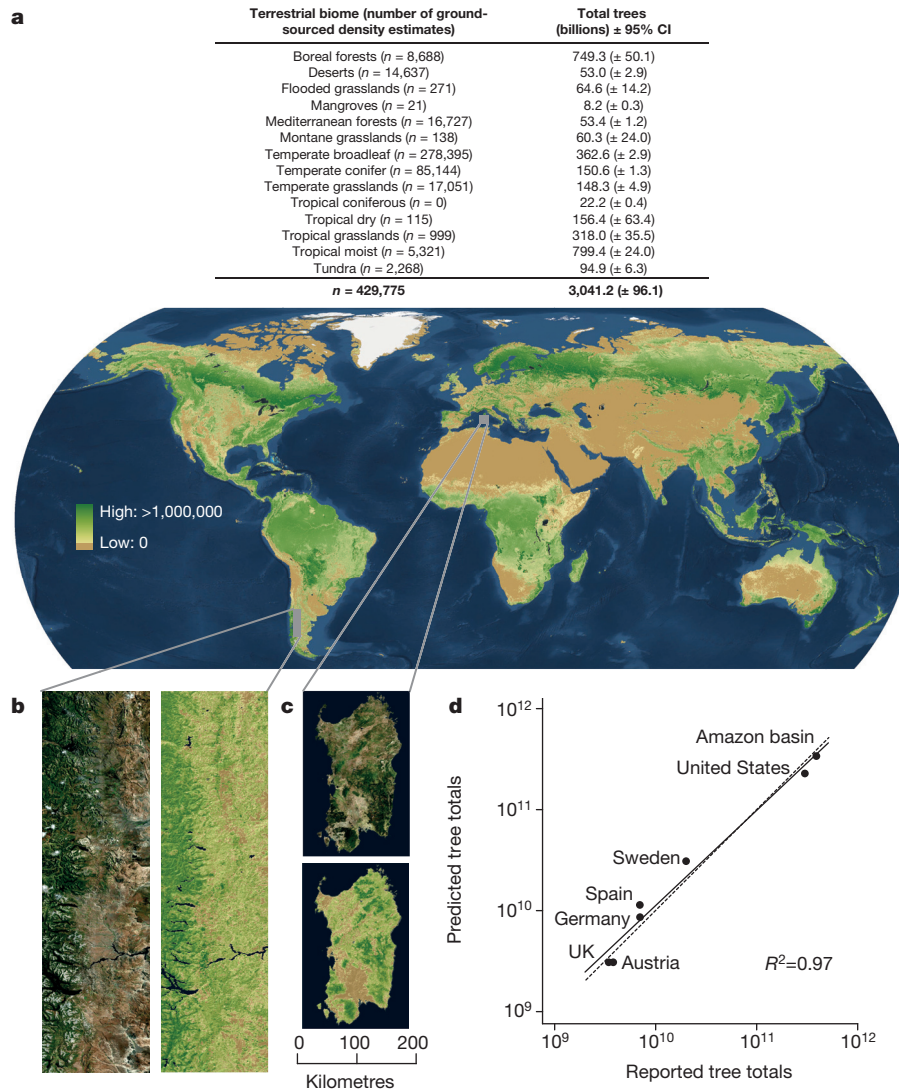
### Within-biome trends

Our models also provide mechanistic insights into potential controls on tree density within biomes (Fig. 5). For example, various climatic parameters correlate with mean forest density within all ecosystem types. Tree density generally increases with temperature (mean annual temperature and temperature seasonality) and moisture availability (precipitation regimes, evapotranspiration or aridity). These patterns are consistent with previous broad-scale tree inventory studies and support the idea that, within ecosystem types, moist, warm conditions are generally optimal for tree growth<sup>11,12</sup>.

Given the generally positive effects of moisture availability and warmth on tree density within biomes, the negative relationships observed in some regions may seem surprising (Fig. 5). This highlights the complex suite of population- and community-level selection pressures that can obscure the expected effects of climate across landscapes. For example, in colder (boreal or tundra) biomes, increasing moisture levels can cause hydric and permafrost conditions in lower lying topographies, which then limit nutrient availability for tree development<sup>20</sup>. In addition, current and historical anthropogenic land use decisions have the potential to drive these relationships in several regions. The negative relationships between tree density and moisture availability in flooded grasslands and tropical dry forests are, for example, likely to be driven by preferential use of moist, productive land for agriculture<sup>21</sup>. As a result, forest ecosystems are often relegated to drier regions, reversing the expected within-biome relationships between moisture availability and tree density. Such effects will vary among countries, depending on human population densities, alternative resource availability and socio-economic status<sup>22,23</sup>.

Along with these indirect effects of human activity, the direct effect of human development (percentage developed and managed land)<sup>6</sup> on tree density represented the only common mechanism across all biomes (Fig. 5). The negative relationships between tree density and anthropogenic land use exemplify how humans contend directly with natural forest ecosystems for space. Whereas the negative effect of human activity on tree numbers is highly apparent at local scales, the present study provides a new measure of the scale of anthropogenic effects, relative to other environmental variables. Current rates of global forest cover loss are approximately  $192,000 \text{ km}^2$  each year<sup>3</sup>. By combining our tree density information with the most recent spatially explicit map of forest cover loss over the past 12 years<sup>3</sup>, we estimate that deforestation, forest management, disturbances and land use change are currently responsible for a gross loss of approximately 15.3 billion trees on an annual basis. Although these rates of forest loss are currently highest in tropical regions<sup>3</sup>, the scale and consistency of this negative human effect across all forested biomes highlights how historical land use decisions have shaped natural ecosystems on a global scale. Using the projected maps of current and historic forest cover provided by the United Nations Environment Programme (<http://geodata.grid.unep.ch>), our map reveals that the global number of trees has fallen





**Figure 4 | The global map of tree density at the 1-km<sup>2</sup> pixel (30 arc-seconds) scale. a**, The scale refers to the number of trees in each pixel. **b**, **c**, We highlight the map predictions for two areas (South American Andes (**b**) and Sardinia (**c**)) and include the corresponding images for visual comparison. All maps and images were generated using ESRI basemap imagery. **d**, A scatterplot as

validation for our broad-scale estimates of total tree number. This shows the relationship between our predicted tree estimates and reported totals for regions with previous broad-scale tree inventories (see Methods for details). The straight line and the dotted line are the predicted best fit line and the 1:1 line, respectively.

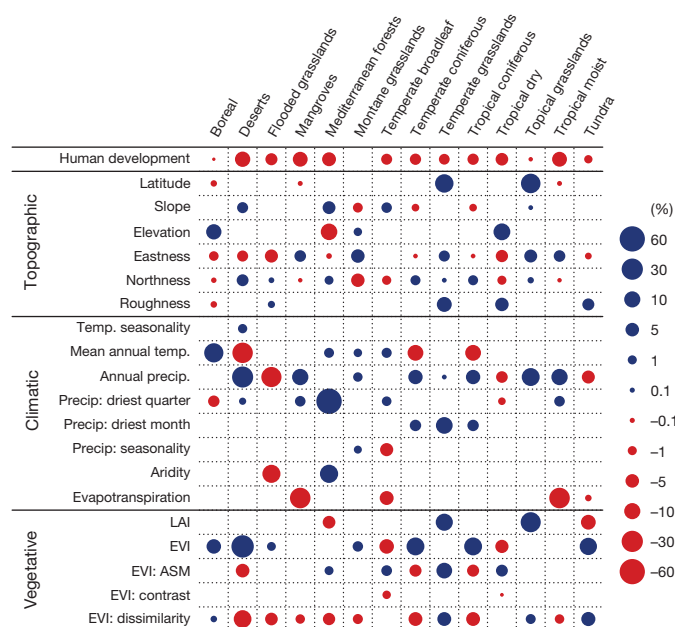
by approximately 45.8% since the onset of human civilization (post-Pleistocene).

## Discussion

The global map of tree density can facilitate ongoing efforts to understand biogeochemical Earth system dynamics<sup>3,6,7,9</sup> by incorporating ecosystem features that relate to elemental cycling rates<sup>9,10</sup>. For example, tree abundance can help to explain some of the variation in carbon storage and productivity within ecosystem types<sup>7,9</sup>, but the strength of these effects remain untested across biomes<sup>8</sup>. We assessed the relationship between tree density and plant carbon storage at a global scale by regressing our plot-level tree counts against modelled estimates of plant biomass carbon in those sites<sup>24</sup>. This revealed a positive effect of tree density on plant carbon storage ( $P < 0.001$ ). However, the strength of the relationship is weak ( $r^2 = 0.14$ ), reflecting the vast array of local ecological forces that can obscure such global trends. For example, the effect of tree density is likely to interact strongly with tree size. Larger trees contain the greatest proportion of carbon in woodlands<sup>25</sup>, but the highest tree densities within a given ecosystem type are often associated with young or recovering forests

characterized by many small trees<sup>13,20</sup>. A thorough understanding of total vegetative carbon storage requires information about both the size and the number of individual trees.

A dense forest environment is a fundamentally different ecosystem from a sparse one and this influences a vast array of biotic and abiotic processes<sup>10–12</sup>. Current remote sensing tools capture some, but not all of this information. The tree density layer that we provide can therefore augment the currently available layers by providing unique insights into ecological dynamics that are not represented by estimates of forest cover or biomass<sup>3,5,6</sup>. It can inform biodiversity estimates and species distribution models by capturing perceivable environmental characteristics that determine habitat suitability for a wide variety of plants and animals<sup>11–13</sup>. Baseline estimates of tree populations are also critical for projecting population- and community-level tree demographics under current and future climate change scenarios<sup>26</sup>, and for guiding local, national, and international reforestation/afforestation efforts<sup>14–17</sup>. Finally, by allowing us to comprehend the global forest extent in terms of tree numbers, this map contributes to our fundamental understanding of the Earth's terrestrial system.



**Figure 5 | Standardized coefficients for the variables included in final biome-level regression models.** Coefficients represent relative per cent change in tree density for one standard deviation increase in the variable. Red and blue circles indicate negative and positive effects on tree density, respectively. Circle size indicates the magnitude of effects. All layers are available at the global scale. Human development = per cent developed and managed land; LAI = leaf area index; EVI = enhanced vegetation index; EVI: ASM = angular second moment of EVI; EVI: contrast = contrast of EVI; and EVI: dissimilarity = dissimilarity of EVI (see Extended Data Table 1).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May; accepted 23 July 2015.

Published online 2 September; corrected online 9 September 2015 (see full-text HTML version for details).

- Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
- Crowther, T. W. *et al.* Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study. *Glob. Change Biol.* **20**, 2983–2994 (2014).
- Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
- Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008).
- Pfeifer, M., Disney, M., Quaife, T. & Marchant, R. Terrestrial ecosystems from space: a review of earth observation products for macroecology applications. *Glob. Ecol. Biogeogr.* **21**, 603–624 (2012).
- Tuanmu, M.-N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* **23**, 1031–1045 (2014).
- Walker, A. P. *et al.* Predicting long-term carbon sequestration in response to CO<sub>2</sub> enrichment: how and why do current ecosystem models differ? *Glob. Biogeochem. Cycles* **29**, 476–495 (2015).
- Asner, G. P. *et al.* A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* **168**, 1147–1160 (2012).
- Fauset, S. *et al.* Hyperdominance in Amazonian forest carbon cycling. *Nature Commun.* **6**, 6857 (2015).
- Slik, J. W. F. *et al.* Environmental correlates of tree biomass, basal area, wood specific gravity and stem density gradients in Borneo's tropical forests. *Glob. Ecol. Biogeogr.* **19**, 50–60 (2010).

- Leathwick, L. R. & Austin, M. P. Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology* **82**, 2560–2573 (2001).
- Oliver, C. D. & Larson, B. C. *Forest Stand Dynamics* (John Wiley & Sons, 1996).
- Riginos, C. & Grace, J. B. Savanna tree density, herbivores, and the herbaceous community: bottom-up vs. top-down effects. *Ecology* **89**, 2228–2238 (2008).
- O'Neil-Dunne, J., MacFaden, S. & Royar, A. A versatile, production-oriented approach to high-resolution tree-canopy mapping in urban and suburban landscapes using GEOBIA and data fusion. *Remote Sens.* **6**, 12837–12865 (2014).
- Guldin, R. W. Forest science and forest policy in the Americas: building bridges to a sustainable future. *For. Policy Econ.* **5**, 329–337 (2003).
- Cao, S. *et al.* Greening China naturally. *Ambio* **40**, 828–831 (2011).
- Oldfield, E. E. *et al.* Growing the urban forest: tree performance in response to biotic and abiotic land management. *Restoration Ecol.* (<http://dx.doi.org/10.1111/rec.12230>) (2015).
- Nadkarni, N. *Between Earth and Sky: Our Intimate Connections to Trees* (Univ. of California Press, 2008).
- ter Steege, H. *et al.* Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092 (2013).
- Bonan, G. B. & Shugart, H. H. Environmental factors and ecological processes in boreal forests. *Annu. Rev. Ecol. Syst.* **20**, 1–28 (1989).
- Meyfroidt, P. & Lambin, E. F. Global forest transition: prospects for an end to deforestation. *Annu. Rev. Environ. Resour.* **36**, 343–371 (2011).
- Rudel, T. K. The national determinants of deforestation in sub-Saharan Africa. *Phil. Trans. R. Soc. Lond. B* **368**, 20120405 (2013).
- Hengeveld, G. M. *et al.* A forest management map of European forests. *Ecol. Soc.* **17**, 53 (2012).
- Kindermann, G. E., McCallum, I., Fritz, S. & Obersteiner, M. A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica* **42**, 387–396 (2008).
- Stephenson, N. L. *et al.* Rate of tree carbon accumulation increases continuously with tree size. *Nature* **507**, 90–93 (2014).
- Zhu, K., Woodall, C. W., Ghosh, S., Gelfand, A. E. & Clark, J. S. Dual impacts of climate change: forest migration and turnover through life history. *Glob. Change Biol.* **20**, 251–264 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank P. Peterkins for her support throughout the study. We also thank Plant for the Planet for initial discussions and for collaboration during the study. The main project was funded by grants to T.W.C. from the Yale Climate and Energy Institute and the British Ecological Society. We acknowledge various sources for tree density measurements and estimates: the Canadian National Forest Inventory (<https://nfi.nfis.org/index.php>), the US Department of Agriculture Forest Service for their National Forest Inventory and Analysis (<http://fia.fs.fed.us/>), the Taiwan Forestry Bureau (which provided the National Vegetation Database of Taiwan), the DFG (German Research Foundation), BMBF (Federal Ministry of Education and Science of Germany), the Floristic and Forest Inventory of Santa Catarina (IFFSC), the National Vegetation Database of South Africa, and the Chilean research grants FONDECYT no. 1151495. For Europe NFI plot data were brought together with input from J. Rondeux and M. Waterinckx, Belgium, T. Bélouard, France, H. Polley, Germany, W. Daamen and H. Schoonderwoerd, Netherlands, S. Tomter, Norway, J. Villanueva and A. Trasobares, Spain, G. Kempe, Sweden. New Zealand Natural Forest plot data were collected by the LUCAS programme for the Ministry for the Environment (New Zealand) and sourced from the National Vegetation Survey Databank (New Zealand) (<http://nvs.landcareresearch.co.nz>). We also acknowledge the BCI forest dynamics research project, which was funded by National Science Foundation grants to S. P. Hubbell, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, numerous private individuals, the Ucross High Plains Stewardship Initiative, and the hard work of hundreds of people from 51 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

**Author Contributions** The study was conceived by T.W.C. and G.H. and designed by T.W.C., K.R.C. and M.A.B. Statistical analyses were conducted by H.B.G., S.M.T., J.R.S., C.B., D.S.M. and T.W.C. and mapping was conducted by H.B.G. and C.B. The manuscript was written by T.W.C. with input from M.A.B., P.C., D.S.M., H.B.G. and C.B., with comments provided by all other authors. Tree density measurements or geospatial data from all over the world were contributed by K.R.C., S.M.T., M.C.D., G.A., M.N.T., W.J., C.Sa., C.St., D.P., T.T., S.G., G.B., S.J.W., S.K.W., M.O.H., G.M.H., G.J.N., E.T., P.B., C.F.L., L.W.P., M.F., A.H., J.H., P.C., A.C.V., P.M.U., S.L.P., C.W.R. and M.S.A.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.W.C. ([thomas.crowther11@gmail.com](mailto:thomas.crowther11@gmail.com)).



## METHODS

**Data collection and standardisation.** Plot-level data were collected from international forestry databases, including the Global Index of Vegetation-Plot Database (GIVD <http://www.givd.info>), the Smithsonian Tropical Research Institute (<http://www.stri.si.edu>), ICP-Level-I plot data which covers most of Europe (<http://www.icp-forests.org>), and National Forest Inventory (NFI) analyses from 21 countries, including the USA (<http://fia.fs.fed.us/>) and Canada (<https://nfi.nfis.org/index.php>). This information was supplemented with data from peer-reviewed studies reporting large international inventories published in the last 10 years (collected using ISI Web of Knowledge, Google Scholar and secondary references)<sup>19,27,28</sup>.

We only included density estimates where individual trees met the criterion of  $\geq 10$  cm diameter at breast height (DBH). Although NFI databases can vary slightly in their definition of a mature tree (for example, the US Forest Service Forest Inventory and Analysis (FIA)<sup>29</sup> defines a tree as a plant with woody stems larger than 12.7 DBH) the vast majority of sources use 10 cm as the DBH cut-off. Indeed, this was the only size class provided by all broad-scale inventories (including the FIA), so density estimates at other DBH values were excluded. This provided a total of 429,775 measurements of forest tree density (each generated at the hectare scale) that were then linked to spatially explicit remote-sensing data and GIS variables to explore the patterns in forest tree density at a global scale. The scale of our plot data (in terms of number and distribution of plots) ensured that any plot location uncertainty or minor changes in global forest area are unlikely to alter mean values or modelled estimates.

**Acquisition and preprocessing of spatial data.** For predictive model development, we selected 20 geospatial covariates from a larger pool of potential covariates based on uniqueness, spatial resolution and ecological relevance (Extended Data Table 1). Covariates were derived through satellite-based remote sensing and ground-based weather stations, and can be loosely grouped into one of four categories: topographic, climatic, vegetative or anthropogenic. Topographic covariates included elevation, slope, aspect (as northness and eastness), latitude (as absolute value of latitude) and a terrain roughness index (TRI). Climatic covariates included annual mean temperature, temperature annual range, annual precipitation, precipitation of driest month, precipitation seasonality (coefficient of variation), precipitation of driest quarter, potential evapotranspiration per hectare per year, and indexed annual aridity. Vegetative covariates included, enhanced vegetation index (EVI), leaf area index (LAI), dissimilarity, contrast, and angular second moment. We also included a single anthropogenic covariate: proportion of urban and/or developed land cover (see Extended Data Table 1).

Several covariates bear special mention. Moving-window analyses were applied to an EVI derived from a multi-year composite of moderate resolution imaging spectroradiometer (MODIS) imagery. From the result, we extracted three second-order textural covariates that reflect the heterogeneity of vegetation, intended to capture difference in vegetative structure. These include angular second moment (the orderliness of EVI among adjacent pixels), contrast (the exponentially weighted difference in EVI between adjacent pixels: see <http://earthenv.org> for details), and dissimilarity (difference in EVI between adjacent pixels). Terrain roughness index (the mean of absolute differences between a cell and its adjacent neighbours) was derived from aggregated Global Multi-Resolution Terrain Elevation Data of 2010. Terrain roughness index was computed using the eight neighbouring pixels, while the others were computed using the four neighbouring pixels located at  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  (see <http://earthenv.org> and ref. 36 for details).

We preprocessed all spatial covariates using ArcMap 10.1 (ESRI, Redlands, CA, 2012) and RStudio 0.97.551 (RStudio, 2012). All covariates were reprojected to the interrupted Goode Homolosine equal-area coordinate system (which maximises spatial precision by amalgamating numerous region-specific equal-area projections) to optimize the areal accuracy of our final figures<sup>30</sup>. These were then resampled to match the coarsest resolution used during analysis (nominal 1 km<sup>2</sup> pixels), and spatially coregistered using nearest neighbour resampling where necessary.

To account for broad-scale differences in vegetation types, we developed spatial models at the biome scale. Individual predictive models were generated within each of 14 broad ecosystem types (delineated by the Nature Conservancy <http://www.nature.org>) to improve the accuracy of estimates.

**Statistical modelling.** We used generalized linear models to generate predictive maps of tree numbers within forested ecosystems for each biome. This approach also enabled us to explore the mechanisms potentially governing patterns in forest tree density within regions (Fig. 5). Due to the inherently interactive nature of climate, soil and human impact factors across the globe, we predicted that there would be pronounced non-independence within the full suite of biophysical variables extracted from the compiled GIS layers. To account for this collinearity, we performed ascendant hierarchical clustering using the *hclustvar* function in

R's *ClustOfVar* package<sup>31</sup> in each biome-level model. This analysis splits the variables into different clusters (similar to principal components) in which all variables correlate with one another. A single best 'indicator' variable is then selected from each cluster, based on squared loading values representing the correlation with the central synthetic variable of each cluster (that is, the first principal component of a PCAmix analysis). This set of 'best' indicator variables for each biome was then included in all subsequent models used to estimate controls on forest tree density.

Using the resulting set of variables, we constructed generalized linear models with a negative binomial error structure (to account for count data that could not extend below zero) for each biome (Extended Data Figs 1, 2 and 3) and performed a multi-model dredging using the dredge function in R's MuMIn package<sup>32</sup>. This function constructs all possible candidate sub-models nested within the global model, identifies the most plausible subset of models for each data set, and then ranks them according to corrected Akaike Information Criterion (AICc) values and AIC likelihood weights (AICw). We derived covariates, coefficients, and variance-covariance matrices for biome-level models through weighted model averaging the dredged model results with cumulative AIC weights at least equal to 0.95 (ref. 33). Given the inherent sampling bias present in our plot data (tree density estimates were only collected in forested ecosystems and non-forested regions are under-represented), our modelling approach was used to generate predictive estimates of forest tree density, and these estimates were subsequently scaled based on the total area of forested land in each pixel (see spatial modelling for details).

**Model validation and testing.** We assessed the model fit by investigating the bias and precision present when predicting mean tree density across an aggregate number of plots. This approach allowed us to test how many plots are required to ensure that the predicted mean (or total) forest density has reasonable bias and precision. 20% of the plots within each biome were randomly omitted before model fitting to serve as an independent data set for model testing. Initial model validation was conducted using the biome-specific regression models (obtained from the remaining 80% of the data) to predict the tree density for each omitted plot. The mean predicted tree density of the omitted data was then regressed against the mean observed tree density of the omitted data for each biome (Fig. 2). In addition, a bootstrapping algorithm was used to quantify the standard deviation of the mean prediction as a function of sample size following ref. 34. For each biome, we generated empirical bootstrap estimates of the standard deviation of the predicted mean using random samples drawn from the withheld validation plots. Specifically, for each biome a bootstrap sample of size  $n$  was selected, with replacement, from the omitted data in that biome. The fitted regression model for that biome (based on the 80% retained data) was used to predict the tree density of each point, and the mean of the  $n$  samples was calculated. This process was repeated 10,000 times for each sample size ( $n = 10, 20, \dots, 500$ ) and in each case the empirical standard deviation of the 10,000 sample mean was calculated and plotted (Fig. 2). Where the number of plot records in a biome fell below the sample size threshold identified through bootstrapping, we used models from the most similar biome available (in terms of phylogenetic relatedness of the dominant tree species and mean tree density from the few plot values collected). This was the case for the two smallest biomes: 'mangroves' (0.23% of land surface) and 'tropical coniferous' (0.46% of land surface) forests, which used models from 'tropical moist' and 'temperate coniferous', respectively.

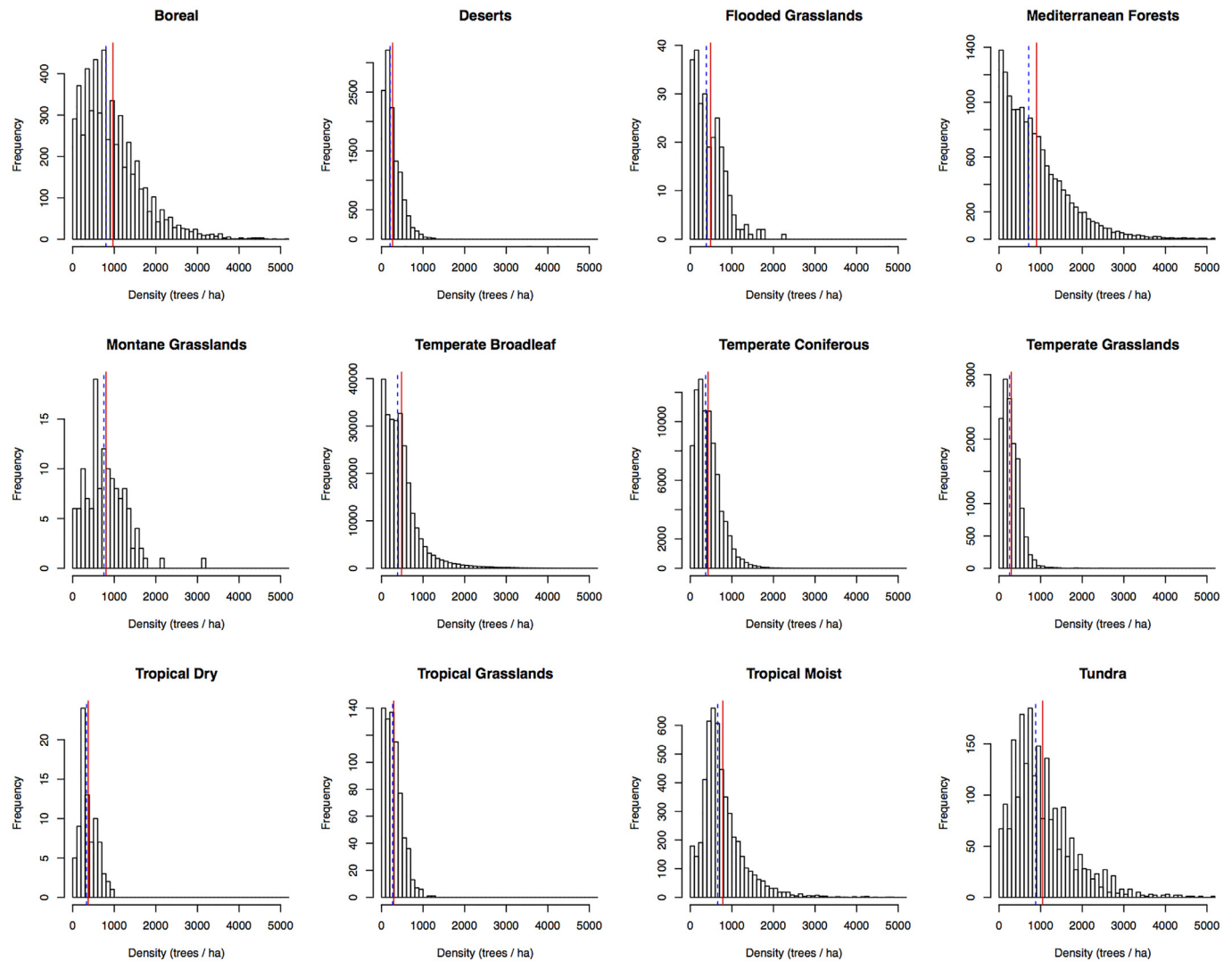
**Spatial modelling.** Following model averaging and bootstrapping, we applied the final negative binomial regression equations used in bootstrapping to pixel-level spatial data at the biome level. Regressions were run in a map algebra framework wherein equation intercepts and coefficients were applied independently to each pixel of our coregistered global covariates to produce a single map of forest tree density on a per-hectare scale. We then scaled our per-hectare forest density estimates to the 1-km<sup>2</sup> scale based on the total area of forested land within each pixel, as estimated by the global 1-km consensus land cover data set for 2014 (ref. 6). This process was then validated using an older (2013) data set that used fine-scale (30 m) forest cover information<sup>3</sup>, which revealed equivalent total tree counts. By multiplying our predicted forest density by the area of forest, we ensured that we did not overestimate tree densities in non-forested sites. From the resulting maps, summary statistics (mean tree density, total tree number) were derived for each polygonal area of interest. The variances of the global and biome-specific totals were calculated using a Taylor series approximation to account for the log-link negative binomial regression function and correlation among the regression-based predicted values<sup>35</sup>.

By generating models at the biome-level, we were able to account for broad-scale differences in vegetation types between biomes, while maintaining high precision of our mean (and total) estimates at the global scale (due to the high

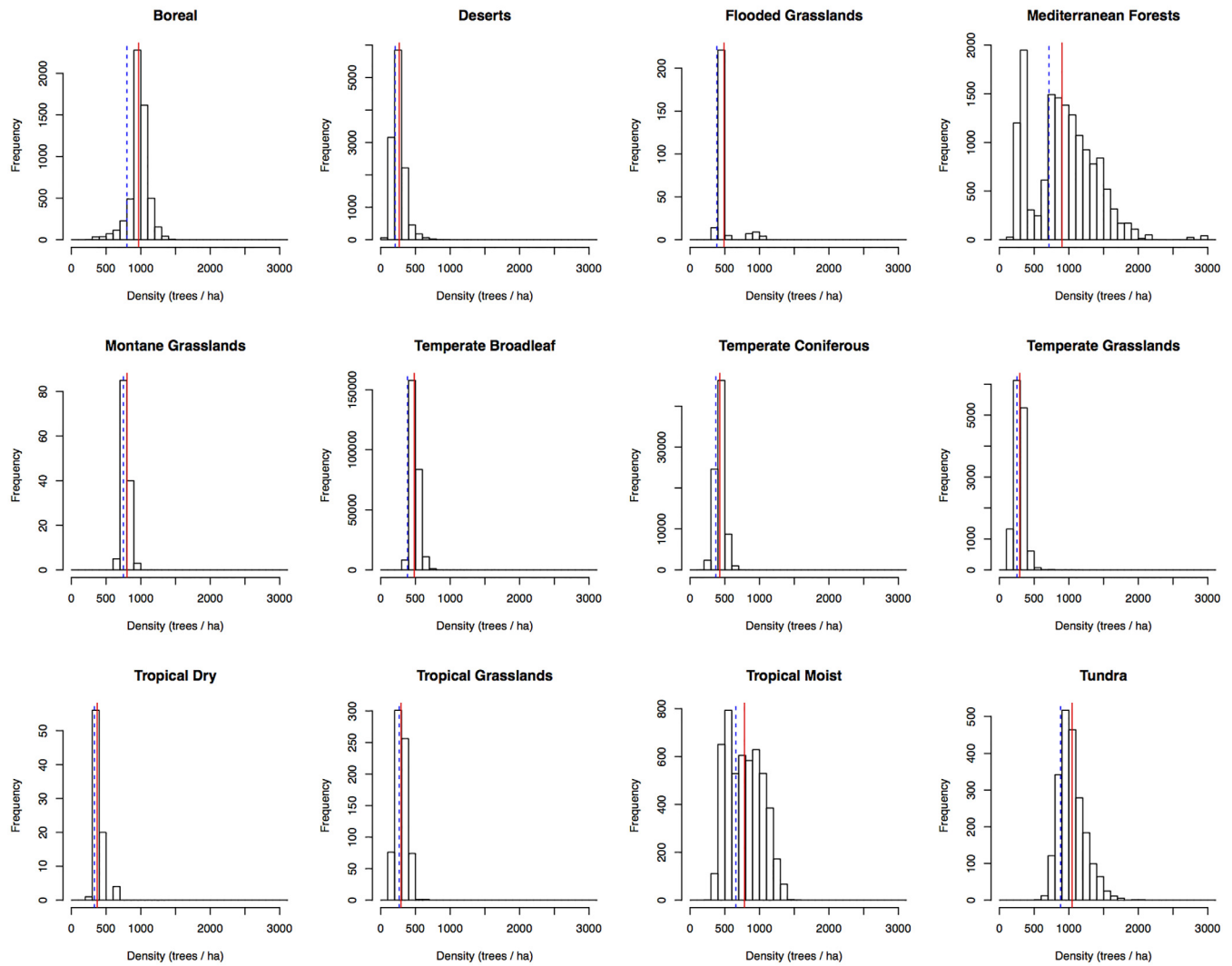
number of plot measurements within biomes). However, biome-level models are limited in their accuracy when predicting tree density at fine-scales, which might ultimately have the potential to alter final numbers. We therefore constructed models within each of 813 global ecoregions (delineated by the Nature Conservancy <http://www.nature.org>) as a validation for the first biome-level approach. We generated models and estimated tree numbers using exactly the same approach as for the biome-level models. Total, and biome-level, tree estimates did not differ significantly ( $P < 0.05$ ) from those generated using the biome-level models (Extended Data Fig. 4).

27. Lewis, S. L. *et al.* Above-ground biomass and structure of 260 African tropical forests. *Phil. Trans. R. Soc. Lond. B* **368**, 20120295 (2013).
28. Brus, D. J. *et al.* Statistical mapping of tree species over Europe. *Eur. J. For. Res.* **131**, 145–157 (2011).
29. USDA Forest Service. Forest Inventory and Analysis National Program <http://fia.fs.fed.us/> (2010).
30. Steinwand, R. S., Hutchinson, J. A. & Snyder, J. P. Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection. *Photogramm. Eng. Remote Sensing* **61**, 1487–1499 (1995).
31. Chavent, M., Kuentz, V., Lique, B. & Saracco, J. ClustOfVar: an R package for the clustering of variables. *J. Stat. Softw.* **50**, 1–16, <http://www.jstatsoft.org/v50/i13/> (2012).
32. Bartoň, K. MuMIn: Model selection and model averaging based on information criteria (AICc and alike). (<https://cran.r-project.org/web/packages/MuMIn/index.html>) (2015).
33. MacKenzie, D. I. *et al.* *Occupancy Estimation and Modeling* (Academic Press, 2005).
34. MacLean, M. G. *et al.* Requirements for labelling forest polygons in an object-based image analysis classification. *Int. J. Remote Sens.* **34**, 2531–2547 (2013).
35. Ståhl, G. *et al.* Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Can. J. For. Res.* **41**, 96–107 (2011).
36. Tuanmu, M.-N. & Jetz, W. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* <http://dx.doi.org/10.1111/geb.12365> (2015).





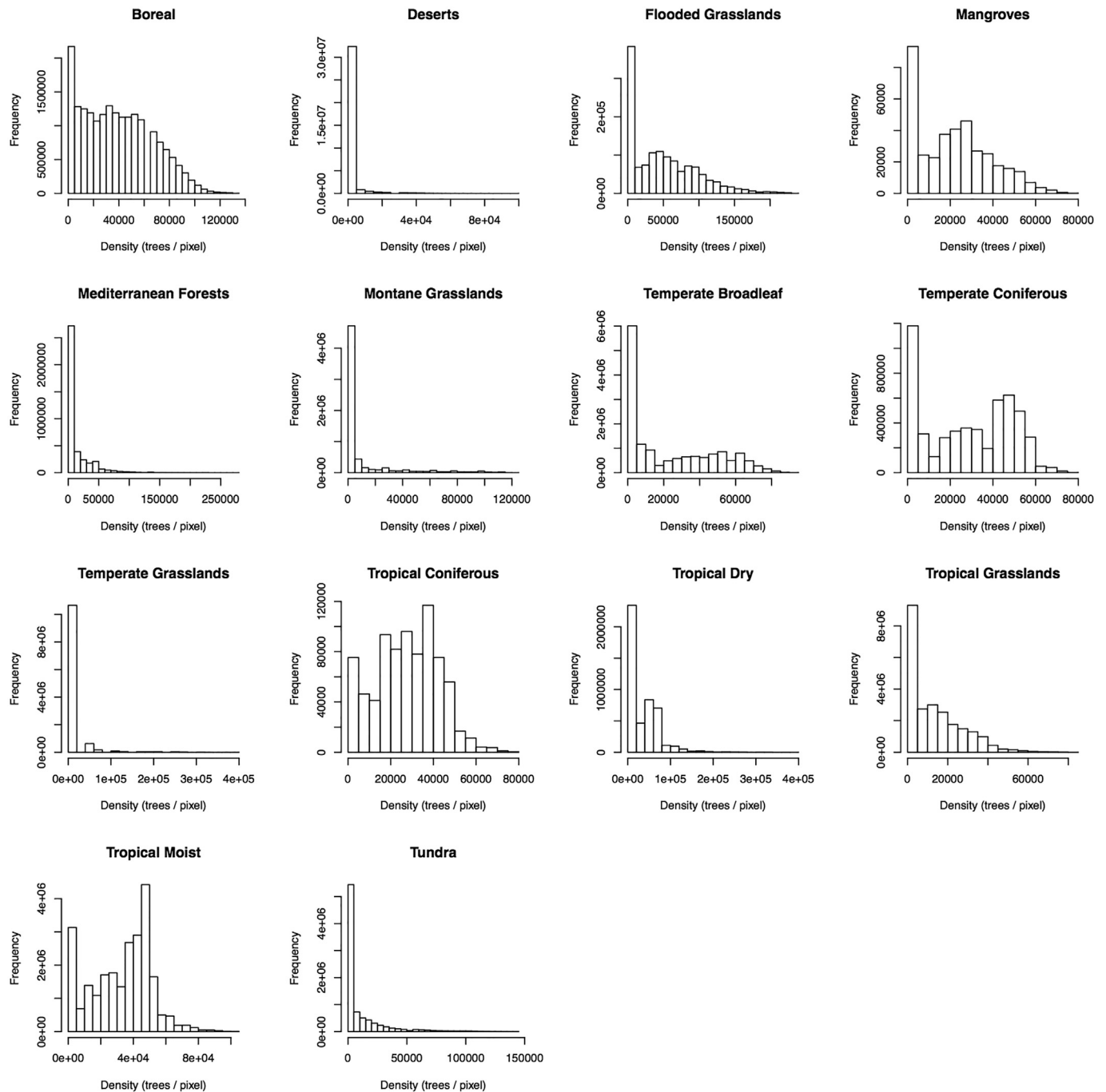
**Extended Data Figure 1 | Histogram of the collected measurements of forest tree density in each biome around the world ( $n = 429,775$ ).** The red line and the blue dotted lines indicate the mean and median for the collected data, respectively. Data in each biome fitted a negative binomial error structure.



**Extended Data Figure 2 | Histogram of the predicted forest tree density values for the locations that density measurements were collected in each biome around the world ( $n = 429,775$ ). The red line and the blue dotted lines**

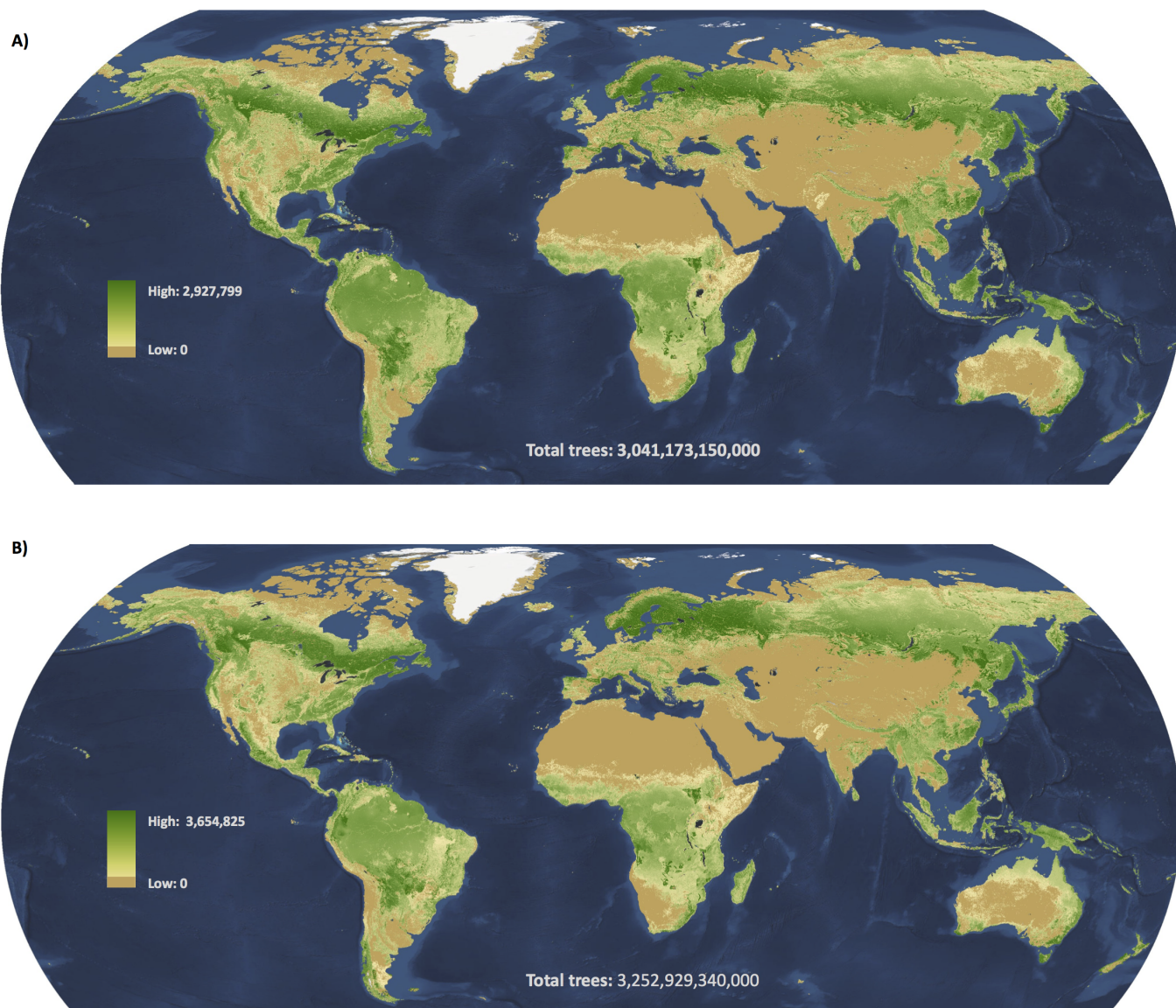
indicate the mean and median for the collected data, respectively. As our models were based on mean values, the majority of points fall on or close to the mean values in each biome.





**Extended Data Figure 3 | Histogram of the total predicted forest tree density values for each pixel within each biome around the world ( $n = 429,775$ ).** This illustrates the spread of pixels throughout each biome, and

highlights that our map accounts for the sampling bias in tree density plots (for example, although we had no zero values in our desert plots, the vast majority of desert pixels contain no trees).



**Extended Data Figure 4 | Comparison between approaches to generate the global tree density map.** The initial map was generated using 14 biome-level models (biomes delineated by The Nature Conservancy <http://www.nature.org>) to account for broad-scale variations in terrestrial vegetation types. With several thousand plot-level density measurements in most biomes, this approach provided highly accurate estimates at the global scale. However, to improve precision at the local scale, we also generated a map using

ecoregion-scale models. Separate models were generated within each of 813 global ecoregions (also delineated by The Nature Conservancy to reflect smaller-scale vegetation types) using exactly the same statistical approach (see Methods). The same 429,775 data points were used to construct each map. Biome-level and ecoregion-level maps provide total tree estimates of 3.041 and 3.253 trillion trees, respectively.



Extended Data Table 1 | Estimates of the total tree number for each of the biomes that contain forested land, as delineated by The Nature Conservancy (<http://www.nature.org>)

Biome	% Total Land Area	n	Total Trees (Billions)	$\pm 2$ SD (Billions)	% Total Trees
Boreal Forests	11.49%	8688	749.34	50.07	24.28%
Deserts	21.01%	14637	52.95	2.92	1.75%
Flooded Grasslands	0.79%	271	64.58	14.19	2.13%
Mangroves	0.23%	21	8.18	0.26	0.27%
Mediterranean Forests	2.43%	16727	53.42	1.20	1.76%
Montane Grasslands	3.88%	138	60.3	24.04	1.99%
Temperate Broadleaf	9.32%	278395	362.6	2.90	11.98%
Temperate Conifer	3.18%	85144	150.57	1.34	4.97%
Temperate Grasslands	7.18%	17051	148.29	4.93	4.90%
Tropical Coniferous	0.48%	0	22.21	0.40	0.73%
Tropical Dry	2.85%	115	156.37	63.42	5.17%
Tropical Grasslands	14.66%	999	318.01	35.52	10.51%
Tropical Moist	14.81%	5321	799.45	23.98	26.41%
Tundra	5.25%	2268	94.89	6.31	3.14%
<b>Total</b>		<b>429775</b>	<b>3041.17</b>	<b>96.07</b>	

# Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth

Jiajun Zhu<sup>1,2,3</sup>, Morgan A. Sammons<sup>1,2</sup>, Greg Donahue<sup>1,2</sup>, Zhixun Dou<sup>1,2</sup>, Masoud Vedadi<sup>4,5</sup>, Matthäus Getlik<sup>6</sup>, Dalia Barsyte-Lovejoy<sup>4</sup>, Rima Al-awar<sup>5,6</sup>, Bryson W. Katona<sup>7</sup>, Ali Shilatifard<sup>8</sup>, Jing Huang<sup>9</sup>, Xianxin Hua<sup>7</sup>, Cheryl H. Arrowsmith<sup>4,10</sup> & Shelley L. Berger<sup>1,2</sup>

**TP53 (which encodes p53 protein) is the most frequently mutated gene among all human cancers. Prevalent p53 missense mutations abrogate its tumour suppressive function and lead to a 'gain-of-function' (GOF) that promotes cancer. Here we show that p53 GOF mutants bind to and upregulate chromatin regulatory genes, including the methyltransferases *MLL1* (also known as *KMT2A*), *MLL2* (also known as *KMT2D*), and acetyltransferase *MOZ* (also known as *KAT6A* or *MYST3*), resulting in genome-wide increases of histone methylation and acetylation. Analysis of The Cancer Genome Atlas shows specific upregulation of *MLL1*, *MLL2*, and *MOZ* in p53 GOF patient-derived tumours, but not in wild-type p53 or p53 null tumours. Cancer cell proliferation is markedly lowered by genetic knockdown of *MLL1* or by pharmacological inhibition of the *MLL1* methyltransferase complex. Our study reveals a novel chromatin mechanism underlying the progression of tumours with GOF p53, and suggests new possibilities for designing combinatorial chromatin-based therapies for treating individual cancers driven by prevalent GOF p53 mutations.**

Most mutant forms of p53 are caused by single amino acid substitutions mapping to the DNA-binding domain<sup>1</sup>. These mutations result in expression of full-length p53 protein, but loss of wild-type (WT) tumour suppressive function<sup>2–4</sup>. The high prevalence of missense substitutions, particularly certain 'hotspot' mutations, suggests a selective advantage during cancer progression. Indeed, these mutants gain neomorphic oncogenic functions, including altered cancer spectrum<sup>2,3</sup>, deregulated metabolic pathways<sup>4,5</sup>, increased metastasis<sup>6,7</sup> and enhanced chemotherapy resistance<sup>8</sup>. Evidence from recent studies points to one potential mechanism of GOF p53, functioning through association with other transcription factors and driving gene transcription in oncogenic pathways, such as the mevalonate pathway<sup>4</sup> and etoposide-resistance pathway<sup>8</sup>. A transcriptional mechanism is further supported by the importance of retaining an intact transactivation domain for oncogenic GOF p53 function<sup>4,9</sup>. Nevertheless, how GOF p53 contributes to major changes of the cancer genome and transcriptome remains to be elucidated<sup>9,10</sup>. Altered chromatin pathways have been implicated in various aspects of cancer<sup>11,12</sup>, given their regulation of genome-wide transcription programs<sup>13,14</sup>. However, to our knowledge, to date there has not been evidence of direct crosstalk between GOF p53 mutants and chromatin regulation.

## Genome-wide binding of GOF p53 mutants

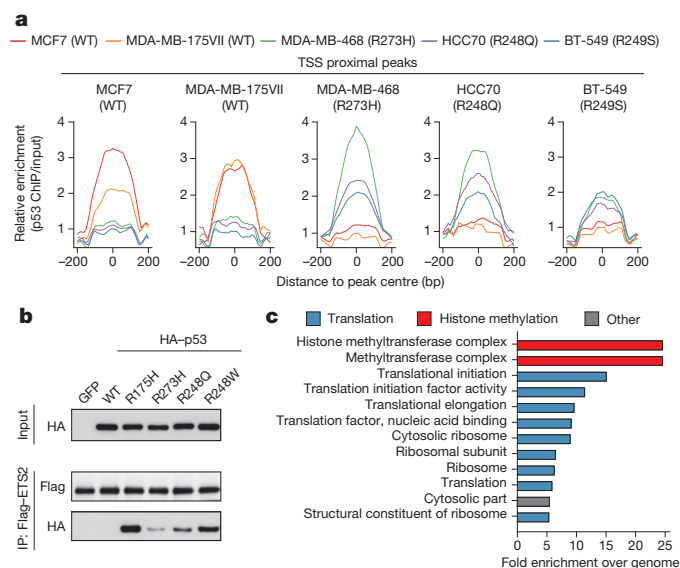
We carried out chromatin immunoprecipitation followed by sequencing (ChIP-seq) to determine genome-wide binding locations of p53 in a panel of breast cancer cell lines: MCF7 (wild-type p53), MDA-MB-175VII (wild-type p53), HCC70 (p53(R248Q)), BT-549 (p53(R249S)), and MDA-MB-468 (p53(R273H)). We found that the binding of p53 to gene-proximal regions (less than 10 kilobases (kb)) of transcription

start sites (TSS) in the two wild-type p53 cell lines strongly resembled each other, whereas these wild-type p53 peaks were highly dissimilar from the peaks in any of the GOF p53 mutants. Notably, p53 binding patterns in the three GOF p53 cell lines were similar when compared to each other (Fig. 1a and Extended Data Fig. 1a). In addition, we aligned published p53(R248W) ChIP-seq data from Li-Fraumeni syndrome (LFS) MDAH087 cells<sup>8</sup>, and again, TSS-proximal peaks of p53(R248W) resembled those of p53(R273H) and p53(R248Q) (Extended Data Fig. 1b, c), but were distinct from the wild-type p53 peaks (Extended Data Fig. 1d, e).

We performed motif analysis for TSS-proximal peaks of the p53(R273H) mutant that predicted the E26 transformation-specific (ETS) motif as the most enriched (Extended Data Fig. 2a), which is distinct from the wild-type p53 motif (Extended Data Fig. 2b). One ETS family member, ETS2, has been shown to consistently associate with mutant p53 (ref. 8). We confirmed that ETS2 interacts with various GOF p53 mutants, but interacts to a much lesser extent with wild-type p53 (Fig. 1b and Extended Data Fig. 2c), as previously noted<sup>8</sup>. Co-immunoprecipitation at endogenous protein levels also demonstrated that ETS2 interacts with GOF p53, but not with wild-type p53 (Extended Data Fig. 2d, e). We analysed ChIP-seq data sets from the ENCODE project for all transcription factors<sup>15,16</sup>, and observed that, compared to other transcription factors, ETS family proteins have significantly higher overlap with GOF p53 TSS-proximal peaks, but not with wild-type p53 TSS-proximal peaks (Extended Data Fig. 2f, g). Notably, in both wild-type and GOF p53 cases, RNA polymerase II (Pol II) group has the highest percentage overlap with p53 peaks, indicative of transcriptional activity. The extent of Pol II overlap is similar to the ETS group in GOF p53 cells, but much higher than the ETS group in wild-type p53 cells (Extended Data Fig. 2f, g).

<sup>1</sup>Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>2</sup>Epigenetics Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>3</sup>Biomedical Graduate Studies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>4</sup>Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada. <sup>5</sup>Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada. <sup>6</sup>Drug Discovery Program, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. <sup>7</sup>Abramson Family Cancer Research Institute, Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. <sup>8</sup>Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, 320 E. Superior Street, Chicago, Illinois 60611, USA. <sup>9</sup>Cancer and Stem Cell Epigenetics, Laboratory of Cancer Biology and Genetics, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA. <sup>10</sup>Princess Margaret Cancer Centre, and Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2C4, Canada.



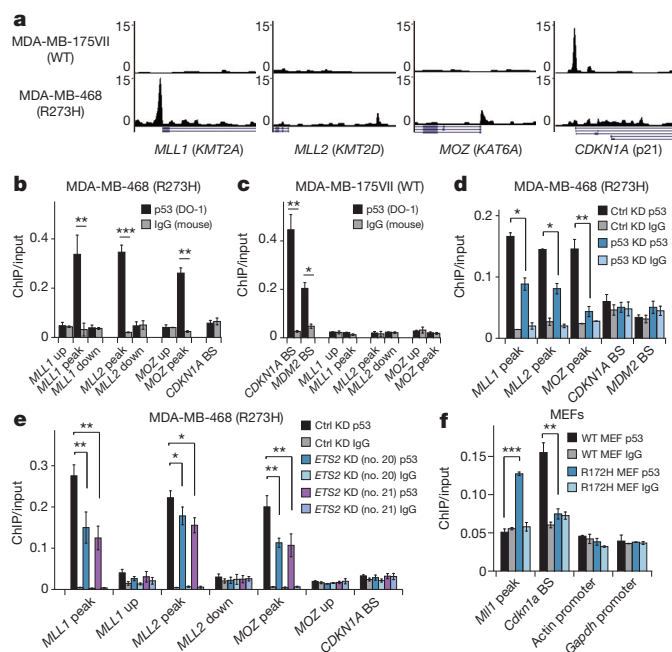


**Figure 1 | Genome-wide binding of GOF p53 mutants.** **a**, Area under the curve analysis showing p53 enrichment (ChIP/input) in five cell lines over TSS-proximal peak regions identified in each cell line. Mann–Whitney *U*-tests were performed to compute significance for combined wild-type (WT) and GOF p53 peaks: MCF7 ( $P = 2.78 \times 10^{-6}$ ), MDA-MB-175VII ( $P = 2.15 \times 10^{-4}$ ), MDA-MB-468 ( $P < 2.2 \times 10^{-16}$ ), HCC70 ( $P = 1.09 \times 10^{-3}$ ), BT-549 ( $P = 3.7 \times 10^{-5}$ ). **b**, Co-immunoprecipitation of HEK293T cell-expressed Flag-ETS2 with *in vitro*-expressed GFP- or HA-tagged p53, followed by western blot. **c**, GO analysis of p53(R273H) TSS-proximal peaks (statistics are shown in Supplementary Table 1). Uncropped blots are in Supplementary Fig. 1.

## GOF p53 targets chromatin regulators

To determine specific functional categories, we performed gene ontology (GO) analysis on TSS-proximal peaks. As expected, DNA damage response pathways were most enriched in wild-type p53 targets (Extended Data Fig. 2h and Supplementary Table 1). In contrast, p53(R273H) bound to genes related to translation and ribosomal synthesis (Fig. 1c and Supplementary Table 1), which was reasonable given the rapid growth rate of these cells. We were particularly intrigued by GOF p53 binding to a group of genes functionally related to histone methylation (Fig. 1c). This was seen in UCSC Genome Browser views at *MLL1* (*KMT2A*) and *MLL2* (*KMT2D*), genes encoding methyltransferases of histone H3 lysine 4 (H3K4) (Fig. 2a) that are components of alternative forms of the COMPASS complex (complex proteins associated with Set1). The other two GOF p53 mutants that we examined, as well as p53(R248W) from LFS MDAH087 cells<sup>8</sup>, all showed similar binding at *MLL1* and *MLL2* (Extended Data Fig. 3a, b, e, f). UCSC Genome Browser views confirmed binding of GOF p53 to a gene encoding a common subunit of COMPASS complexes, *RBBP5* (Extended Data Fig. 3h). In contrast, wild-type p53 did not appear to bind any of these genes, although as expected it bound promoter regions of its canonical targets, including *CDKN1A* (which encodes p21 protein), *MDM2* and *BBC3* (also known as *PUMA*) (Fig. 2a and Extended Data Fig. 3c, i, j). We then analysed a large set of 600 chromatin regulators for potential GOF p53 binding, and found an additional group of chromatin regulatory genes that showed peak enrichment (Supplementary Table 2). Of particular interest among these was *MOZ* (*KAT6A*), a histone acetyltransferase, and UCSC Genome Browser views confirmed the presence of GOF p53 but not wild-type p53 (Fig. 2a and Extended Data Fig. 3d, g).

Using ChIP–quantitative PCR (ChIP–qPCR), we validated the binding of GOF p53 to *MLL1*, *MLL2*, and *MOZ* genes, but not adjacently upstream or downstream of the peak regions (Fig. 2b and Extended Data Fig. 4a–c). Moreover, we confirmed GOF p53 binding to all other targets in the ‘histone methylation’ GO category (*RBBP5*, *OGT* and *PPP1CC*), and to a few additional chromatin factors



**Figure 2 | GOF p53 mutants directly target chromatin regulators.** **a**, UCSC Genome Browser views of p53 occupancy over promoter regions of *MLL1*, *MLL2*, *MOZ* and *CDKN1A*. **b**, **c**, ChIP–qPCR showing p53 or IgG enrichment (ChIP/input) in MDA-MB-468 (**b**) and MDA-MB-175VII (**c**) cells. BS, p53 binding site. Up, upstream of GOF p53 binding region; peak, GOF p53 binding region; down, downstream of GOF p53 binding region. A schematic of amplicon locations is shown in Extended Data Fig. 4a. **d**, **e**, ChIP–qPCR showing p53 enrichment changes upon reduction of p53 (**d**) or ETS2 (**e**) by shRNA-mediated knockdown. Numbers 20 and 21 denote two short hairpins, the sequences of which are shown in Supplementary Table 3. **f**, ChIP–qPCR showing p53 or IgG enrichment in MEFs bearing wild-type p53 or p53(R172H). Error bars represent mean  $\pm$  s.e.m.;  $n = 3$ ; two-tailed Student's *t*-test: \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

(including *SMARCD2* and *DCAF10*), in all three GOF p53 cell lines (Extended Data Fig. 4d–f). We verified the ChIP–qPCR results with a second p53 polyclonal antibody, FL393 (Extended Data Fig. 4g). In parallel experiments with both p53 antibodies, wild-type p53 showed binding to the *CDKN1A* and *MDM2* canonical binding sites, but not to any of the GOF p53 targets tested (Fig. 2c and Extended Data Fig. 4h). We also examined a pancreatic cancer cell line, PANC-1 (p53(R273H)), and observed a similar binding pattern (Extended Data Fig. 4i), suggesting a general phenomenon in various cancer types. Furthermore, the ChIP–qPCR signal of GOF p53 was attenuated upon p53 knockdown (Fig. 2d). Knockdown of *ETS2* also led to reduced binding of GOF p53 over *MLL1* and *MOZ*, and to a lesser extent, over the *MLL2* peak region (Fig. 2e). To test the association of GOF p53 near *MLL1* in a non-tumour background, we performed ChIP–qPCR in primary mouse embryonic fibroblasts (MEFs) with GOF p53 or wild-type p53, and consistently, mouse GOF p53 showed significant enrichment over the *MLL1* promoter region (Fig. 2f).

## GOF p53 regulates MLL, MOZ, and histone modifications

To examine whether GOF p53 is required for expression of the chromatin regulators, we reduced GOF p53 levels in human cancer cells and found that the mRNA levels of *MLL1*, *MLL2* and *MOZ* were also decreased (Fig. 3a and Extended Data Fig. 5a); whereas no change was detected when the level of wild-type p53 was reduced (Extended Data Fig. 5b). Simply increasing wild-type p53 protein levels by stabilization mediated by the compound nutlin did not recapitulate activation of the chromatin regulators (Extended Data Fig. 5c, d). *MLL1* protein levels were also decreased in the GOF p53 knockdown (Fig. 3b), but not by wild-type p53 knockdown (Extended Data Fig. 5e), as was also observed for *MOZ* protein levels (Extended Data

Fig. 5f). Reduction of ETS2 levels led to decreased expression of *MLL1* and *MOZ*, and to a lesser extent, *MLL2* (Fig. 3c and Extended Data Fig. 5g), which was in accordance with the relative binding changes of GOF p53 to these genes (Fig. 2e). We verified the *ETS2* knockdown result in another GOF p53 cell line, BT-549, and detected similarly decreased expression of *MLL1* and *MOZ*, and to a lesser extent, *MLL2* (Extended Data Fig. 5h, i). We performed Pol II ChIP-qPCR and observed concomitantly decreased Pol II occupancy specifically over *MLL1*, *MLL2*, and *MOZ* TSS regions upon *ETS2* knockdown (Fig. 3d). We examined the importance of another ETS family member, ETS1. By contrast, *ETS1* knockdown had no effect on the expression of *MLL1*, *MLL2* or *MOZ* (Extended Data Fig. 5j, k), nor did it alter GOF p53 or Pol II binding (Extended Data Fig. 5l, m). As ETS family proteins consist of 28 members<sup>17</sup>, it is likely that additional ETS protein(s) other than ETS1 may be involved. Nevertheless, our observations are consistent with previous studies showing that ETS2, but not ETS1, is important in mediating GOF p53 function<sup>8,18</sup>.

The regulation of histone-modifying enzymes led to investigation of the cognate histone post-translational modifications (PTMs). We observed a global decrease in histone H3 lysine 9 acetylation (H3K9ac, catalysed by MOZ<sup>19</sup>) in response to knockdown of GOF p53, whereas other histone acetylation marks did not show notable changes (Fig. 3e and Extended Data Fig. 5n, o, p). The reduction of H3K9ac was also observed when the level of MOZ itself was decreased by short hairpin RNA (shRNA) (Extended Data Fig. 5q). In contrast, H3K4 tri-methylation and H3K4 mono-methylation (H3K4me3 and H3K4me1, catalysed by MLL1 and MLL2, respectively<sup>20</sup>) showed only a slight global reduction upon GOF p53 knockdown (Fig. 3e and Extended Data Fig. 5n, o, p). This is reasonable, however, given that H3K4 is methylated by six members of the COMPASS complexes<sup>20</sup>, and previous studies showed that inhibiting or knocking one of them out did not substantially change global H3K4 methylation<sup>21,22</sup>.

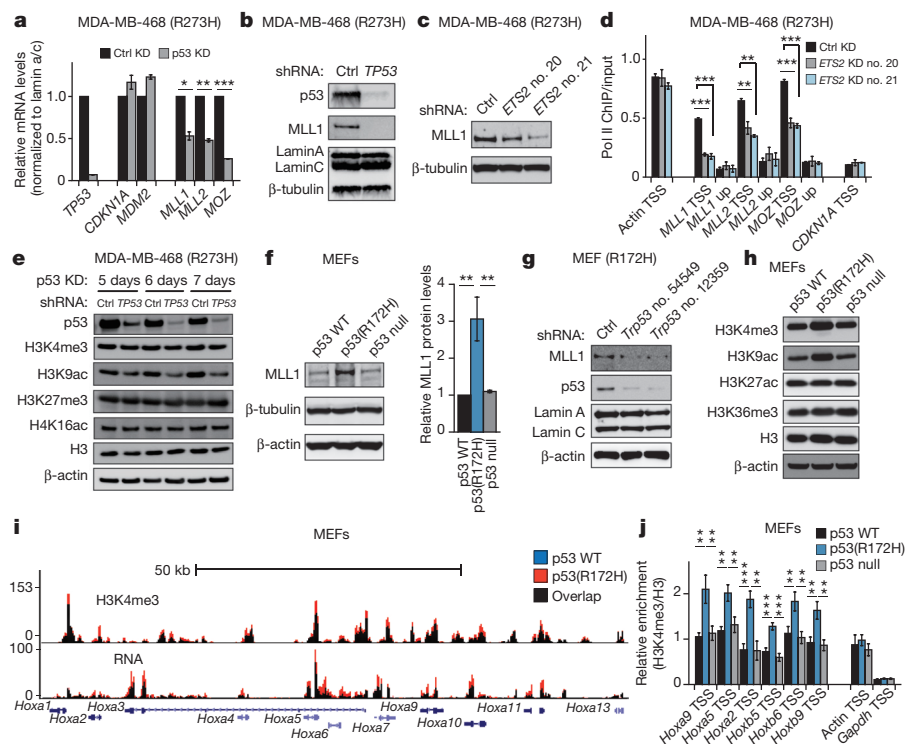
We further validated the regulation of *MLL1*, *MLL2*, and *MOZ* by GOF p53 in the knock-in MEFs. We found significantly higher expression of these genes in GOF p53 MEFs than in wild-type p53 MEFs or in MEFs derived from p53 (gene *Trp53*) knockout mice (p53 null MEFs) (Fig. 3f and Extended Data Fig. 6a–c). Furthermore, when GOF p53 was reduced, *MLL1* expression was also lowered (Fig. 3g and Extended

Data Fig. 6d), and ectopically expressing GOF p53 in p53 null MEFs enhanced *MLL1* expression (Extended Data Fig. 6e, f). GOF p53 MEFs also showed higher global level of H3K9ac, and a slight elevation of H3K4me3, compared with wild-type p53 or p53 null MEFs (Fig. 3h and Extended Data Fig. 6g). Notably, other histone modifications associated with active gene transcription, including H3K27ac and H3K36me3, remained at comparable levels (Fig. 3h). In addition, H3K4me3 or H3K9ac did not change upon knockdown of wild-type p53 (Extended Data Fig. 6h), even though cell growth was increased as expected (Extended Data Fig. 6i). Together, these data suggest that changes in H3K4me3 and H3K9ac are specific to GOF p53 directly activating MLL1 and MOZ enzymes.

The modest global change in H3K4me3 in the presence of GOF p53 prompted investigation of local changes in H3K4 methylation. We performed RNA-seq and H3K4me3 ChIP-seq in MEFs with endogenous wild-type p53 or GOF p53. Compared with the genome-wide average, known MLL1 target genes<sup>22</sup> were more highly expressed and displayed higher H3K4me3 enrichment in GOF p53 MEFs (Extended Data Fig. 6j). For example, we observed increased H3K4me3 level and RNA expression within the *Hoxa* gene cluster (Fig. 3i and Extended Data Fig. 6k), a well-studied target of MLL1 and commonly upregulated in leukaemia<sup>23,24</sup>. Conversely, wild-type p53 targets, such as *Cdkn1a*, showed decreased RNA expression and TSS-associated H3K4me3 in GOF p53 MEFs (Extended Data Fig. 6l). Notably, H3K4me3 enrichment at the TSS of genes in GOF p53 MEFs was slightly, but significantly higher at a genome-wide level than in wild-type p53 MEFs (Extended Data Fig. 6m), consistent with the slight global increase of H3K4me3 (Fig. 3h). We validated the H3K4me3 ChIP-seq and RNA-seq results by ChIP-qPCR and RT-qPCR, respectively, observing significantly higher H3K4me3 enrichment in GOF p53 MEFs, and higher expression of *Hox* genes, than in wild-type p53 or p53 null MEFs (Fig. 3j and Extended Data Fig. 6n).

### MLL1 is essential for cancer phenotype of GOF p53

Previous studies have revealed that cells expressing GOF p53 rely on it for cell growth and survival<sup>12,26</sup>. GOF p53 knockdown in cancer cells led to a strong decrease in cell proliferation (Extended Data Fig. 7a). By contrast, lowering of wild-type p53 levels resulted in elevated



**Figure 3 | GOF p53 mutants regulate MLL and MOZ and thereby histone post-translational modifications.** **a**, RT-qPCR measuring mRNA level changes upon shRNA-mediated p53 knockdown. **b**, **c**, Western blot of MLL1 protein level changes upon p53 (**b**) or ETS2 (**c**) knockdown. **d**, ChIP-qPCR showing Pol II enrichment changes upon shRNA-mediated ETS2 knockdown. **e**, Western blot showing histone modification changes upon p53 knockdown over time. **f**, Western blot (left) and quantification (right) of endogenous MLL1 protein levels in MEFs with wild-type p53, p53(R172H) or p53 null. **g**, Western blot indicating MLL1 level changes upon p53 knockdown in MEFs with p53(R172H). Numbers 54549 and 12359 denote two short hairpins, sequences of which are shown in Supplementary Table 3. **h**, Western blot measuring endogenous histone modification levels in MEFs bearing wild-type p53, p53(R172H) or p53 null. **i**, Overlaid UCSC Genome Browser views of H3K4me3 and RNA levels in MEFs with p53 WT or R172H, over *Hoxa* gene cluster. **j**, ChIP-qPCR showing H3K4me3 enrichment (H3K4me3 ChIP/H3 ChIP) in MEFs with wild-type p53, p53(R172H) or p53 null. Uncropped blots are in Supplementary Fig. 1. Error bars represent mean  $\pm$  s.e.m.;  $n = 3$ ; two-tailed Student's *t*-test; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .



growth (Extended Data Figs 6i and 7b). To investigate the function of GOF p53 driving chromatin regulators, we carried out the same time course, and found that the reduction of MLL1 or MLL2 in GOF p53 cancer cells led to a striking loss of cell growth (Extended Data Fig. 7c), phenocopying the knockdown of GOF p53 itself. By contrast, knockdown of *MLL1* or *MLL2* had a minimal effect on wild-type p53 cancer cells (Extended Data Fig. 7d, e).

We addressed the importance of this pathway to tumour-relevant phenotypes, first by examining the ability of cancer cells to form colonies. Reduction of MLL1 led to a decreased colony formation ability of MDA-MB-468 cells (p53(R273H)) (Fig. 4a and Extended Data Fig. 7f), but had little effect on the colony formation efficiency of MCF7 cells (wild-type p53) (Fig. 4b and Extended Data Fig. 7g). Similar results were observed in breast cancer cells BT-549 (p53(R249S)) and pancreatic cancer cells PANC-1 (p53(R273H)) (Extended Data Fig. 7h, i). We further confirmed the tumour formation phenotype in anchorage-independent growth assays in soft agar, showing that decreasing MLL1 specifically reduced the growth and colony size of GOF p53 cancer cells, but not wild-type p53 cancer cells (Extended Data Fig. 7j, k). We also investigated tumour growth on NOD-scid-gamma (NSG) immunodeficient mice. Knockdown of *MLL1* led to strongly reduced tumour formation ability in GOF p53 cells, as compared to GOF p53 cells with a non-targeting scrambled control knockdown (Fig. 4c, e). In contrast, *MLL1* knockdown did not alter the tumour formation ability of wild-type p53 cancer cells (Fig. 4d, e), again supporting a specific role for MLL1 in cancers with GOF p53, but not wild-type p53.

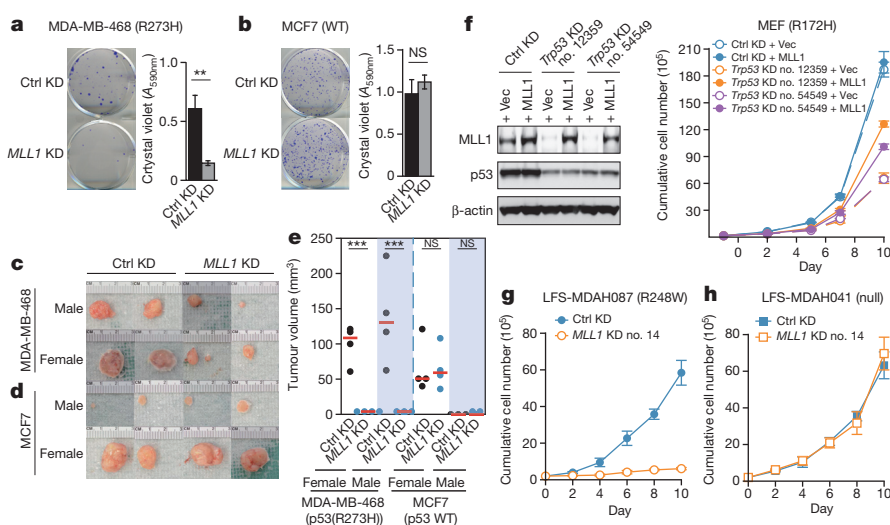
To further explore a critical role that these chromatin regulators may play in supporting growth of GOF p53 cells, and to rule out possible confounding factors in established cancer cell lines, we performed *MLL1* knockdown in the primary MEFs with knock-in GOF p53. Consistently, MLL1 reduction resulted in decreased proliferation of GOF p53 MEFs (Extended Data Fig. 8a). Importantly, re-expression of MLL1 in GOF p53 MEFs with p53 knockdown partially rescued the growth defects (Fig. 4f); partial rescue probably results from GOF p53 driving expression of multiple downstream targets, as described above. These results strongly indicate a direct role of MLL1, functioning

downstream of GOF p53 in maintaining proliferation of GOF p53 cells. We also performed *MLL1* knockdown (Extended Data Fig. 8b) in human non-cancer LFS cells—MDAH087 (p53(R248W)) and MDAH041 (p53 null<sup>27</sup>; Extended Data Fig. 8c). Similar to the results obtained in cancer cells and in MEFs, *MLL1* knockdown reduced the growth rate of GOF p53 LFS cells (Fig. 4g and Extended Data Fig. 8d), again phenocopying the knockdown of GOF p53 itself (Extended Data Fig. 8e), but did not reduce the growth of either p53 null LFS cells (Fig. 4h and Extended Data Fig. 8f), nor primary non-cancer cells with wild-type p53 (IMR90 lung fibroblasts, Extended Data Fig. 8g, h). Re-expression of MLL1 again partially rescued the growth reduction by GOF p53 knockdown in LFS cells (Extended Data Fig. 8i). In addition, *MLL2* knockdown also decreased GOF p53 LFS cell proliferation, but not p53 null LFS cells (Extended Data Fig. 8j, k).

## COMPASS inhibitors reduce GOF p53 cell growth

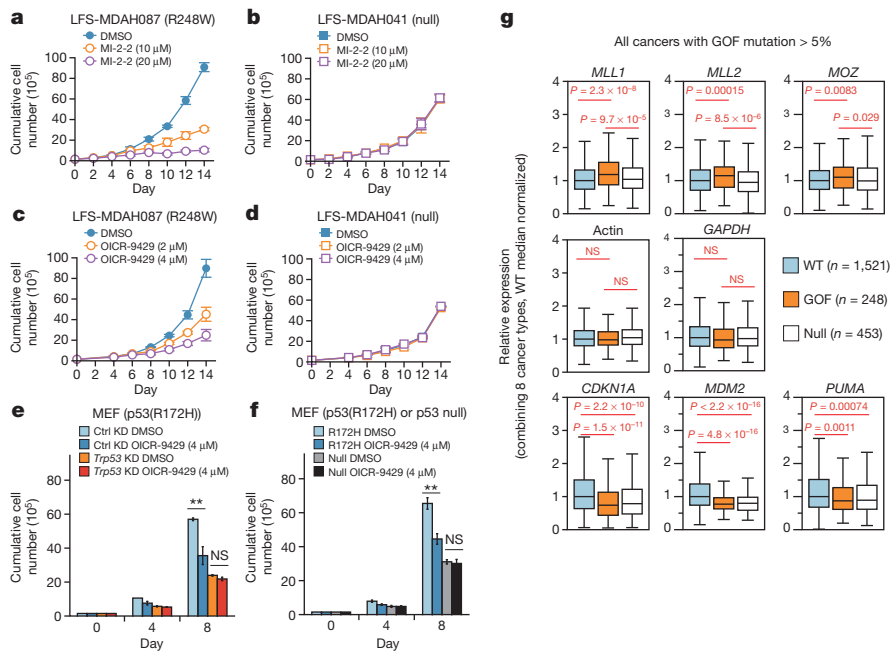
Chromatin regulators have emerged as promising targets of small molecule compounds in various human diseases including cancer<sup>11,28</sup>. Menin is a scaffold protein of the COMPASS complex<sup>20</sup>, directly interacting with the amino-terminal of MLL1<sup>29–31</sup>, and is crucial for MLL1 activity and for maintenance of a subtype of leukaemia<sup>32,33</sup>. We treated both GOF p53 and p53 null LFS cells with the previously reported menin antagonist, MI-2-2 (refs 34, 35). In agreement with the *MLL1* genetic knockdown experiments, MI-2-2 showed a dose-dependent inhibition of GOF p53 cell growth (Fig. 5a), but had very little effect on p53 null cells (Fig. 5b).

Recently, inhibition of MLL1 function has also been demonstrated by targeting its interaction with the WDR5 subunit of the COMPASS complex<sup>36,37</sup>. As a second approach to pharmacological inhibition of MLL1 activity, we used OICR-9429, a newly characterized antagonist of interaction of WDR5 with MLL1 (ref. 38). This non-peptide, drug-like molecule binds to WDR5 in the MLL1 binding site of WDR5 ( $K_d = 93 \pm 28$  nM), and disrupts the assembly of the WDR5/MLL1/RbBP5 complex in cells with half-maximum inhibitory concentration ( $IC_{50}$ ) values below  $1 \mu\text{M}$ <sup>38</sup>. In striking similarity to MI-2-2, we found a dose-dependent inhibition by OICR-9429 of GOF p53 LFS cell growth



**Figure 4 | *MLL1* knockdown reduces the proliferation and cancer phenotype of GOF p53 cells.** **a, b,** Colony formation (left) and quantification (right) in MDA-MB-468 (**a**) or MCF7 (**b**) cells with non-targeting control (ctrl) or *MLL1* knockdown (KD). Two-tailed Student's *t*-test; \*\* $P < 0.01$ ; NS,  $P > 0.05$ ;  $n = 3$ ; the other two biological replicates are shown in Extended Data Fig. 7f, g. **c, d,** Excised xenograft tumours 20 weeks after NSG immunodeficient mice were subcutaneously injected with MDA-MB-468 (**c**) or MCF7 (**d**) cells carrying control or *MLL1* knockdown. Two representative images out of four total in each group are shown. **e,** Xenograft tumour volumes measured 10 weeks after initial injection described in **c** and **d**. Palpable tumours

at a size below  $4 \text{ mm}^3$  were recorded as  $4 \text{ mm}^3$  due to difficulties in measurement. Zeros indicate that the mouse did not have a palpable tumour. Red horizontal lines shown as average tumour volume of all four mice in each group. Mann-Whitney *U*-test; \*\*\* $P < 0.001$ ; NS,  $P > 0.05$ . **f,** Growth curve analysis and corresponding western blot in p53(R172H) MEFs with control or p53 knockdown, and vector control (Vec) or MLL1 overexpression. **g, h,** Growth curve analysis in LFS MDAH087 (**g**) or MDAH041 (**h**) cells with control or *MLL1* knockdown. Uncropped blots shown in Supplementary Fig. 1. Error bars represent mean  $\pm$  s.e.m.;  $n = 3$ .



**Figure 5 | COMPASS inhibitors specifically reduce GOF p53 cell growth.** **a, b**, Growth curve analysis of LFS MDAH087 (**a**) and MDAH041 (**b**) cells treated with DMSO, and 10 μM or 20 μM MI-2-2. **c, d**, Growth curve analysis of LFS MDAH087 (**c**) and MDAH041 (**d**) cells treated with DMSO, and 2 μM or 4 μM OICR-9429. **e**, Growth analysis of p53(R172H) MEFs carrying control or p53 knockdown, treated with DMSO or 4 μM OICR-9429. **f**, Growth analysis of p53(R172H) or p53 null MEFs treated with DMSO or 4 μM OICR-9429. **g**, Box plots of TCGA RNA expression profiles in tumours with wild-type p53, p53 GOF or p53 null. Mann-Whitney *U*-tests were performed to compute significance; NS,  $P > 0.05$ .

(Fig. 5c), and, again, little effect on p53 null LFS cells (Fig. 5d). Moreover, in the genetically controlled MEF cells, we observed similar results, that OICR-9429 specifically inhibits cell proliferation of GOF p53 MEFs (Fig. 5e, f), but not when GOF p53 is reduced (Fig. 5e) or in p53 null MEFs (Fig. 5f). These results provide strong evidence for a specific growth inhibitory effect of pharmacological drugs in targeting MLL COMPASS complex activity downstream of GOF p53.

We examined the significance of our findings in the context of human tumour samples, by analysing The Cancer Genome Atlas (TCGA). Based on p53 mutational status, we grouped tumour samples into: (1) wild type (no detectable p53 mutation); (2) GOF (missense mutation of R175H, R248Q, R248W, R249S or R273H); or (3) p53 null (p53 nonsense mutations or frameshift truncations). Tumours with other types of p53 mutations (other missense mutations, inframe insertion/deletion, or splicing mutations) were not included in further analysis, due to an unpredictable effect on the downstream chromatin regulators. We also focused our analysis on cancer types that include more than 5% of samples in the group comprising GOF p53. We first combined all samples from these cancer types, and observed significantly higher RNA expression of *MLL1*, *MLL2* and *MOZ* in GOF p53 tumours, compared to either wild-type p53 or p53 null tumours (Fig. 5g, top panels). As controls, expression levels of housekeeping genes including actin (*ACTB*) and *GAPDH* are consistent across the three groups (Fig. 5g, middle panels), whereas expression levels of wild-type p53 targets *CDKN1A*, *MDM2* and *PUMA* are significantly higher in the wild-type p53 group than the GOF p53 or p53 null group (Fig. 5g, lower panels). Next, we examined individual cancer types and observed similar gene expression patterns as the combination of all cancers (Extended Data Fig. 9a–f). Notably, given the heterogeneous population of tumour samples, and the small sample size of certain groups, not all pairwise comparisons are statistically significant, although the same trends always hold that GOF p53 tumours express higher levels of *MLL1*, *MLL2*, and *MOZ* than the other two groups. This is also true with canonical wild-type p53 targets, that is, although not all comparisons are statistically significant, the wild-type p53 groups always show higher levels of *CDKN1A*, *MDM2* and *PUMA* than the GOF p53 or p53 null tumour groups.

## Discussion

Our results indicate that distinct prevalent GOF p53 mutants bind to a common newly identified group of gene targets genome-wide, to drive

expression of genes comprising a chromatin signature. The GOF p53 mapping occurs immediately associated with ETS motifs, and GOF p53 binds directly to ETS2, indicating that the substitutions in the DNA-binding domain of p53 unleash a latent interaction with ETS family transcription factors, as previously suggested<sup>6</sup>. Within this chromatin signature gene group targeted by GOF p53, the COMPASS methyltransferase pathway appears to be particularly well represented, but the new binding includes other chromatin regulators, such as the acetyltransferase *MOZ*. We find that expression of these modifying enzymes is dependent on GOF p53, which in turn elevates activating histone modifications, including H3K4me3 and H3K9ac. Our evidence points to MLL downstream pathways as key targets of GOF p53. Thus, as is the case in leukaemia bearing translocations of MLL, MLL pathways may contribute to GOF p53 oncogenic phenotypes and therefore cancer progression.

Importantly, our findings in both human cancer cells and LFS cells show that GOF p53 cells lose growth and tumour formation potential with similar timing kinetics upon knockdown of *MLL1* as they do with knockdown of GOF p53. A key comparison—to cancer and LFS cells that express wild-type p53 or are null for p53—shows very little response to *MLL1* knockdown. Hence, GOF p53 cells appear particularly dependent for growth on the MLL1 pathway. We provide further evidence of GOF p53 cell growth dependence on the COMPASS complex, by analysing cell sensitivity to two different pharmacological small compound inhibitors. These compounds target menin or WDR5 interaction with MLL1, and inhibit proliferation of LFS cells and MEFs expressing GOF p53 but not p53 null. The effects of the inhibitors are thus analogous to direct knockdown of *MLL1*. Hence, we conclude that a large cohort of GOF-p53-driven cancers, the growth of which was not previously known to be dependent on chromatin pathways, may be amenable to epigenetic therapeutics.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 June 2014; accepted 29 July 2015.

Published online 2 September 2015.

- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Lang, G. A. *et al.* Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* **119**, 861–872 (2004).



3. Olive, K. P. *et al.* Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
4. Freed-Pastor, W. A. *et al.* Mutant p53 disrupts mammary tissue architecture via the mevalonate pathway. *Cell* **148**, 244–258 (2012).
5. Zhang, C. *et al.* Tumour-associated mutant p53 drives the Warburg effect. *Nat. Commun.* **4**, 2935 (2013).
6. Subramanian, M. *et al.* A mutant p53/let-7i-axis-regulated gene network drives cell migration, invasion and metastasis. *Oncogene* **34**, 1094–1104 (2015).
7. Weissmueller, S. *et al.* Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor  $\beta$  signaling. *Cell* **157**, 382–394 (2014).
8. Do, P. M. *et al.* Mutant p53 cooperates with ETS2 to promote etoposide resistance. *Genes Dev.* **26**, 830–845 (2012).
9. Scian, M. J. *et al.* Modulation of gene expression by tumor-derived p53 mutants. *Cancer Res.* **64**, 7447–7454 (2004).
10. Garritano, S., Inga, A., Gemignani, F. & Landi, S. More targets, more pathways and more clues for mutant p53. *Oncogenesis* **2**, e54 (2013).
11. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12–27 (2012).
12. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Med.* **19**, 1438–1449 (2013).
13. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
14. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
15. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
16. Gertz, J. *et al.* Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol. Cell* **52**, 25–36 (2013).
17. Hollenhorst, P. C., McIntosh, L. P. & Graves, B. J. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* **80**, 437–471 (2011).
18. Xiong, S. *et al.* Pla2g16 phospholipase mediates gain-of-function activities of mutant p53. *Proc. Natl Acad. Sci. USA* **111**, 11145–11150 (2014).
19. Voss, A. K., Collin, C., Dixon, M. P. & Thomas, T. Moz and retinoic acid coordinately regulate H3K9 acetylation, Hox gene expression, and segment identity. *Dev. Cell* **17**, 674–686 (2009).
20. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
21. Cao, F. *et al.* Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol. Cell* **53**, 247–261 (2014).
22. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell Biol.* **29**, 6074–6085 (2009).
23. Milne, T. A. *et al.* MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol. Cell* **10**, 1107–1117 (2002).
24. Nakamura, T. *et al.* ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol. Cell* **10**, 1119–1128 (2002).
25. Lim, L. Y., Vidovic, N., Ellisen, L. W. & Leong, C. O. Mutant p53 mediates survival of breast cancer cells. *Br. J. Cancer* **101**, 1606–1612 (2009).
26. Alexandrova, E. M. *et al.* Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment. *Nature* **532**, 352–356 (2015).
27. Zhu, Q., Wani, G., Wani, M. A. & Wani, A. A. Human homologue of yeast Rad23 protein A interacts with p300/cyclic AMP-responsive element binding (CREB)-binding protein to down-regulate transcriptional activity of p53. *Cancer Res.* **61**, 64–70 (2001).
28. Dawson, M. A., Kouzarides, T. & Huntly, B. J. Targeting epigenetic readers in cancer. *N. Engl. J. Med.* **367**, 647–657 (2012).
29. Huang, J. *et al.* The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature* **482**, 542–546 (2012).
30. Yokoyama, A. *et al.* Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell Biol.* **24**, 5639–5649 (2004).
31. Caslini, C. *et al.* Interaction of MLL amino terminal sequences with menin is required for transformation. *Cancer Res.* **67**, 7275–7283 (2007).
32. Thiel, A. T., Huang, J., Lei, M. & Hua, X. Menin as a hub controlling mixed lineage leukemia. *Bioessays* **34**, 771–780 (2012).
33. Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207–218 (2005).
34. Grembecka, J. *et al.* Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. *Nature Chem. Biol.* **8**, 277–284 (2012).
35. Shi, A. *et al.* Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood* **120**, 4461–4469 (2012).
36. Karatas, H. *et al.* High-affinity, small-molecule peptidomimetic inhibitors of MLL1/WD5 protein-protein interaction. *J. Am. Chem. Soc.* **135**, 669–682 (2013).
37. Karatas, H., Townsend, E. C., Bernard, D., Dou, Y. & Wang, S. Analysis of the binding of mixed lineage leukemia 1 (MLL1) and histone 3 peptides to WD repeat domain 5 (WDR5) for the design of inhibitors of the MLL1–WDR5 interaction. *J. Med. Chem.* **53**, 5179–5185 (2010).
38. Grebien, F. *et al.* Pharmacological targeting of the Wdr5-MLL interaction in C/EBP $\alpha$  N-terminal leukemia. *Nature Chem. Biol.* **11**, 571–579 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Tainsky for the LFS cell lines; A. Weller, J. Glover and the Stem Cell and Xenograft Core at the University of Pennsylvania for help with the tumour xenograft experiments. S.L.B. is supported by NIH grant R01 CA078831. M.A.S. is supported by a Postdoctoral Fellowship from the American Cancer Society. X.H. is supported in part by a pilot grant from ITMAT of the University of Pennsylvania. A.S. is supported by NIH grant R01 GM069905. The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from AbbVie, Bayer, Boehringer Ingelheim, Genome Canada through the Ontario Genomics Institute (OGI-055), GlaxoSmithKline, Janssen, Lilly Canada, Merck, the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, Takeda, and the Wellcome Trust (092809/Z/10/Z). Funding was also provided to C.H.A. from the Canadian Cancer Society Research Institute.

**Author Contributions** J.Z. and S.L.B. initiated and led the project. J.Z., M.A.S., Z.D. and S.L.B. designed the experiments and interpreted the data. J.Z. and M.A.S. performed the experiments. J.Z., M.A.S. and G.D. analysed all next-generation sequencing data. M.V., M.G., D.B.-L., R.A. and C.H.A. developed OICR9429. B.W.K. and X.H. contributed to the use of menin inhibitor. A.S. and J.H. contributed to reagents used in this study. J.Z., M.A.S., C.H.A. and S.L.B. composed the manuscript. All authors reviewed and commented on the manuscript.

**Author Information** ChIP-seq and RNA-seq data can be accessed through NCBI Gene Expression Omnibus (GEO) database under accession number GSE59176. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L.B. ([bergers@upenn.edu](mailto:bergers@upenn.edu)).

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Cell culture.** MCF7, MDA-MB-175VII, HCC70, BT-549, and MDA-MB-468 cell lines were obtained from American Type Culture Collection (ATCC), and were cultured in a 37 °C incubator at 20% oxygen, in standard tissue culture medium (DMEM with 10% FBS, 100 units per ml penicillin and 100 µg per ml streptomycin) supplied with non-essential amino acids. Li-Fraumeni Syndrome cell lines MDAH087 and MDAH041 were obtained from Michael A. Tainsky (Wayne State University, Detroit, MI) as a gift, and were cultured in a 37 °C incubator at 3% oxygen, in standard tissue culture medium. R172H knock-in mice were generated by Tyler Jacks (Massachusetts Institute of Technology)<sup>3</sup> and obtained from the NCI Mouse Repository. Primary MEFs from 13.5-day embryos were generated as previously described<sup>39</sup>, and cultured in standard tissue culture medium in a 37 °C incubator at 3% oxygen condition.

**Western blot and antibodies.** Cells were lysed in modified RIPA buffer containing 150 mM NaCl, 1% NP-40, 50 mM Tris-Cl, pH 8.0, and 1% SDS, supplemented with protease inhibitors (Life Technologies, number 78446) before use. Protein concentration was determined by BCA protein assay (Life Technologies, number 23227), following which equal amount of proteins were loaded and separated in polyacrylamide gels. Proteins were then transferred to nitrocellulose membrane. Antibodies used in this study were as follows: p53 monoclonal antibody DO-1 (Calbiochem EMD); p53 polyclonal antibody FL393 (Santa Cruz Biotechnology, sc-6243). Flag (Sigma, M2, F1804), HA (Rockland, 600-401-384), histone H3 (abcam, ab1791), H3K4me1 (abcam, ab8895), H3K4me2 (Active Motif, 39142), H3K4me3 (abcam, ab8580), H3K9ac (Active Motif, 39137), H3K14ac (Active Motif, 39616), H3K27ac (abcam, ab4729), H3K36me3 (abcam, ab9050), ETS2 (Santa Cruz Biotechnology, sc-351), MLL1 (Bethyl Laboratories, A300-086A), MOZ (Novus Biologicals, 21620002), mouse p53 antibody for ChIP experiments (Santa Cruz Biotechnology, sc-1312 (M-19)), mouse p53 antibody for western blot analysis (Cell Signaling Technology, number 2524), RNA polymerase II (abcam, ab817).

**Co-immunoprecipitation.** Flag tagged ETS2 protein was transfected (Life Technologies, number 11668019) and expressed in HEK293T cells and then subjected to immunoprecipitation with Flag antibody conjugated protein G Dynabeads (Life Technologies, number 10004D). Following stringent washes, HA-tagged wild-type p53 or GOF p53 (generated by *in vitro* translation (Thermo, number 88881)) was added to co-immunoprecipitate with Flag-ETS2 in buffer containing: 20 mM Tris, pH 8.0, 137 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 1% NP-40, and protease inhibitors. Endogenous co-immunoprecipitation experiments were performed in buffer containing: 20 mM Tris, pH 8.0, 137 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 1% NP-40, 10% glycerol, with protease and phosphatase inhibitors, and 12.5 U ml<sup>-1</sup> benzonase (Novagen, 70746).

**Bacterial expression and GST pull-down.** GST-tagged ETS2 constructs were transformed and expressed in BL21-CodonPlus *E. coli*. Bacterial lysates were incubated with glutathione beads (Life Technologies, number G2879) at 4 °C for 2 h, and washed four times with buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton, 1 mM DTT, supplemented with 100 µM PMSF. The *in vitro* translated (Thermo, number 88881) HA-tagged wild-type p53 or GOF p53 proteins were pre-cleared with GST at 4 °C for 1 h and the resulting supernatant was subjected to GST pull-down with GST or GST-ETS2. The product was then washed and subjected to western blot analysis.

**RT-qPCR, ChIP-qPCR and ChIP-sequencing.** RNA was isolated from cells using RNeasy kit (Qiagen, number 74106). RNA was then reverse transcribed to cDNA (Life Technologies, number 4387406), then qPCR was performed for quantification using standard procedures on a 7900HT Fast-Real-Time PCR platform (ABI). ChIP was performed as previously described<sup>40</sup>, with modifications. In brief, cells were crosslinked in 1% formaldehyde (Thermo, number 28906) in PBS for 10 min at room temperature. After glycine quenching, cell pellets were collected and lysed as previously described<sup>40</sup>, and then subjected to sonication with the Covaris sonicator (S220). The supernatant was then diluted in the same sonication buffer but without N-lauroylsarcosine, and subjected to immunoprecipitation with corresponding antibodies at 4 °C overnight. The beads were then washed and DNA was reverse-crosslinked and purified. Following ChIP, DNA was quantified by qPCR using standard procedures on a 7900HT Fast-Real-Time PCR platform (ABI), or sequencing libraries were prepared using NEBNext Ultra library preparation procedure, and then sequenced on Illumina Hi-Seq platform at the Next-Generation Sequence Core at University of Pennsylvania, or on Illumina Next-Seq platform in the Epigenetics Program at the University of Pennsylvania. All primer sequences used in this study are available in Supplementary Table 3.

**Growth curve measurement.** 200,000 cells were seeded on 950 mm<sup>2</sup> surface area (one well of 6-well plate) on day 0. Cell number was measured every two days with

Countess automated cell counter (Life Technologies) following standard procedures and default parameter settings, after which 200,000 cells were plated back for the next count. For shRNA-mediated knockdown experiments, cells were seeded 7 days after the initial infection of shRNA-containing lentivirus, during which puromycin selection was completed and cells were returned to normal growth medium. For small compound inhibitor treatment experiments, inhibitors or DMSO vehicle control were added on day 0 as cells were seeded, and refreshed every other day as cells were counted and replated. All short hairpin sequences used in this study are available in Supplementary Table 3.

**Colony-formation assay.** After lentiviral infection of shRNA constructs and puromycin selection, 2,000 cells were seeded per well in 6-well plates. After three weeks, cell colonies were fixed with 1% paraformaldehyde and stained with 0.1% crystal violet (for 15 min). For quantification, the crystal violet dye was released into 10% acetic acid and measured at A<sub>590nm</sub> (OD590).

**Soft agar anchorage-independent growth assay.** The base layer of soft agar contained complete DMEM media (10% FBS, 100 units per ml penicillin and 100 µg per ml streptomycin) with 1% agar; the top layer of soft agar contained complete DMEM media with 0.7% agarose and was mixed with 5,000 cells and plated over the base layer. Colonies were fixed and stained with 0.005% crystal violet (for 1 h), and visible colonies were counted.

**Tumour xenograft assay.** A total of four male and four female mice (*Mus musculus*, strain NOD.Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ, Jackson Labs (stock number 005557)) between the ages of 38 and 45 days old were used per treatment for tumour xenograft experiments. All animal experiments described adhere to policies and practices approved by the University of Pennsylvania Institutional Biosafety Committee (IBC) and the Institutional Animal Care and Use Committee (IACUC). Cells were collected after shRNA (MLL1 or non-targeting control) mediated knockdowns. Then 1.5 million cells were injected subcutaneously per mouse. Tumour size was measured by calipers 10 weeks after subcutaneous injection. Tumour size was measured in two dimensions, and tumour volume was calculated as 0.5 × length × width<sup>2</sup>. All mice were euthanized 20 weeks after subcutaneous injection. Tumours were then excised and photographed.

**ChIP-sequencing and RNA-sequencing analysis.** Human cell sequencing reads were aligned to human genome hg18 using Bowtie2 (ref. 41). For p53 ChIP-seq, significant regions of enrichment (peaks) were called using HOMER (Salk Institute, <http://homer.salk.edu>). For area under the curve analysis, ChIP-seq tags from each cell line were counted at TSS proximal peaks (200 bp around peak centres) of every cell line (including itself) as indicated. Heat maps of p53 enrichment across a 5 kb region (±2.5 kb from peak centre, bin = 10) in MCF7, MDA-MB-175VII, MDA-MB-468, HCC70, BT-549 cell lines were generated using HOMER and visualized using JavaTreeView. Sequencing reads from MEFs ChIP-seq experiments were aligned to the mouse reference genome mm9 using Bowtie2. Strand-specific mouse RNA-seq experiments were aligned to the mm9 reference genome and reference transcriptome. FPKM expression values were counted for each exon and merged into a single gene model using HOMER.

**Motif analysis.** To determine associated sequence motifs for wild-type p53 or GOF p53 peaks, all TSS proximal peaks (filtered to remove peaks overlapping with satellite DNA) were pared down to the central 50 bp and used as input to MEME and the SeqPos utility in Cistrome (central 100 bp as required by SeqPos). MEME was instructed to search for the top 10 motifs appearing 0 or more times in each sequence, and SeqPos was run with default parameters.

**Gene ontology analysis.** GO terms associated with wild-type p53 or GOF p53 binding sites were determined in the following way. ChIP-seq TSS proximal peaks were associated with the nearest ENSEMBL transcript and processed using DAVID. The FDR was controlled at 1% and GO terms with fewer than 5 associated transcripts or a fold-enrichment over the genomic background under fivefold were discarded.

**Intersection with ENCODE transcription factor data sets.** Transcription factor peak coordinates (hg18 assembly) were obtained from the ENCODE project repository (<http://www.encodeproject.org>) in BED format. TSS proximal p53 ChIP-seq peak regions were intersected with all transcription factor binding-site data using BEDTools, with overlap inferred if a minimum of a single base pair was in common.

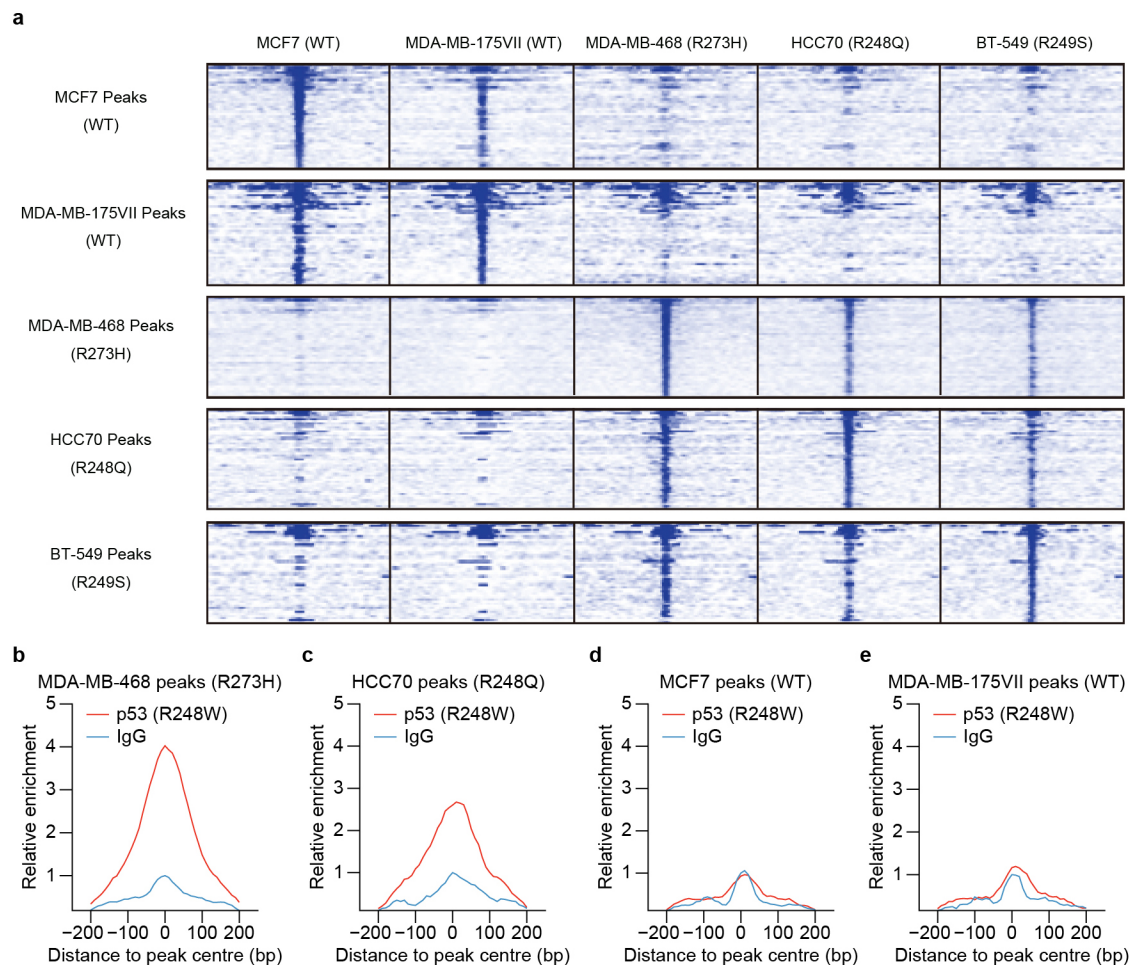
**TCGA analysis.** Exome sequencing and RNA sequencing data sets were obtained from TCGA (<https://tcga-data.nci.nih.gov/tcga/>). Based on p53 mutational status from the exome sequencing data sets, we grouped tumour samples into: (1) wild type (tumours without detectable p53 mutation); (2) GOF (tumours with p53 single missense mutation of R175H, R248Q, R248W, R249S or R273H); and (3) null (tumours with p53 nonsense mutations or frameshift truncations). Tumours with other types of p53 mutations (other missense mutations, inframe insertion/deletion, or splicing mutations) were not included in further analysis, due to an unpredictable effect on the downstream chromatin regulators. Cancer types that include more than 5% samples in group 2 were included for the combined analysis, in which RNA expression values were normalized to the wild-type group



median. For individual cancer type analysis, original RNA expression values (normalized read counts or RPKM values) from TCGA data sets were used.

**OICR-9429.** OICR-9429 was developed using structure-guided medicinal chemistry and peptide displacement assays starting from ‘Compound 3’ previously reported in ref. 42, as part of the Chemical Probe Program of the Structural Genomics Consortium. OICR-9429 is highly specific for WDR5 and was shown to have >100-fold selectivity over 300 other chromatin ‘reader’ domains, methyl-transferases, and other non-epigenetic targets. The details of its structure, discovery and characterization are described in ref. 38.

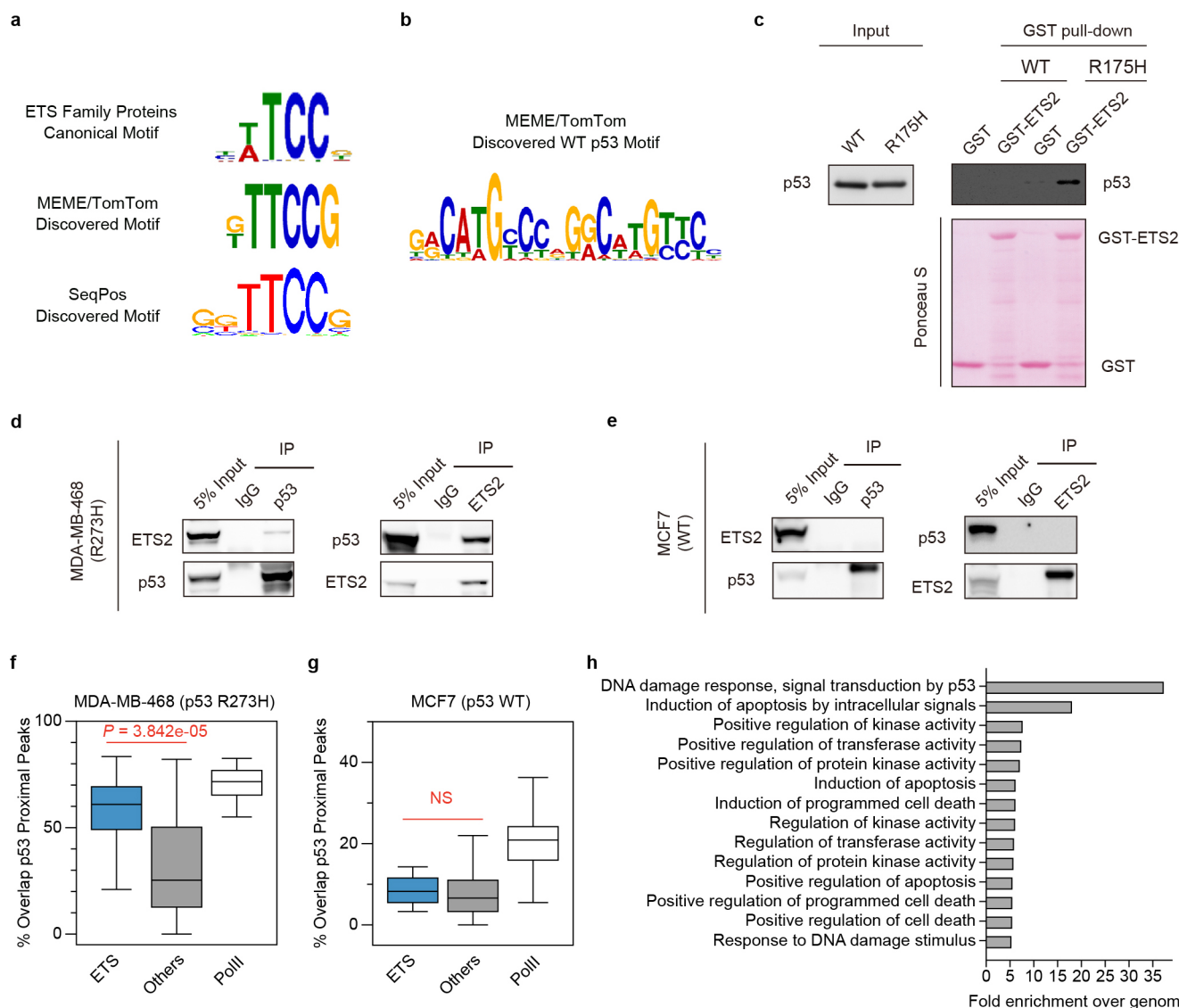
39. Lee, K. H. *et al.* A genomewide study identifies the Wnt signaling pathway as a major target of p53 in murine embryonic stem cells. *Proc. Natl Acad. Sci. USA* **107**, 69–74 (2010).
40. Shah, P. P. *et al.* Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* **27**, 1787–1799 (2013).
41. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
42. Senisterra, G. *et al.* Small-molecule inhibition of MLL activity by disruption of its interaction with WDR5. *Biochem. J.* **449**, 151–159 (2013).



**Extended Data Figure 1 | Distinct GOF p53 mutants have similar genome-wide binding patterns, but are different from that of wild-type p53.** **a**, Heat maps showing the enrichment of p53 peaks ( $\pm 2,500$  bp around peak centre) identified from each cell line (rows) in all five cell lines (columns) examined by

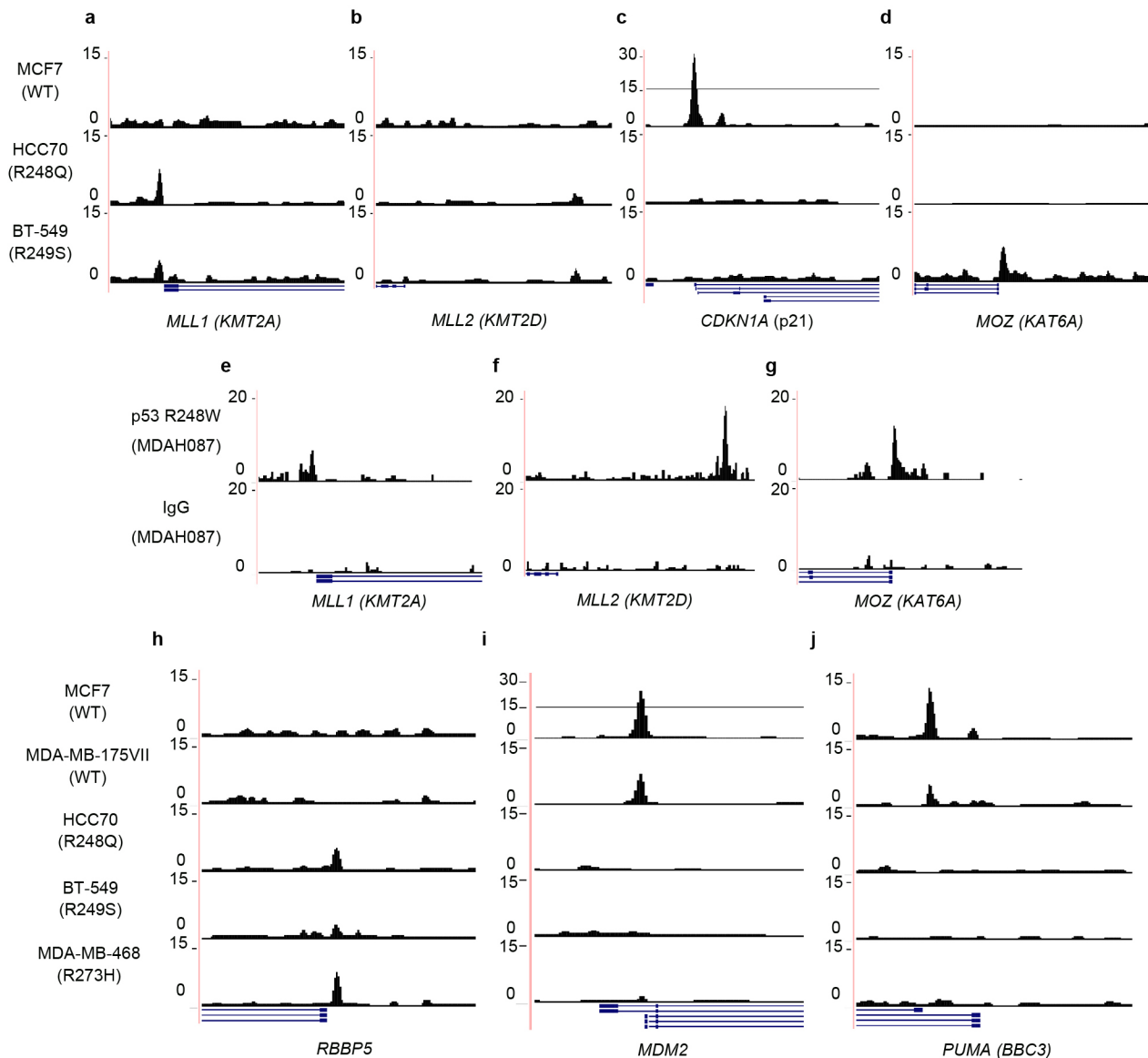
ChIP-seq. **b–e**, Area under the curve, meta-peak analysis showing GOF p53(R248W) or IgG ChIP-seq signal enrichment from MDAH087 cells over TSS-proximal peaks identified in MDA-MB-468 (**b**), HCC70 (**c**), MCF7 (**d**) and MDA-MB-175VII (**e**) cells.





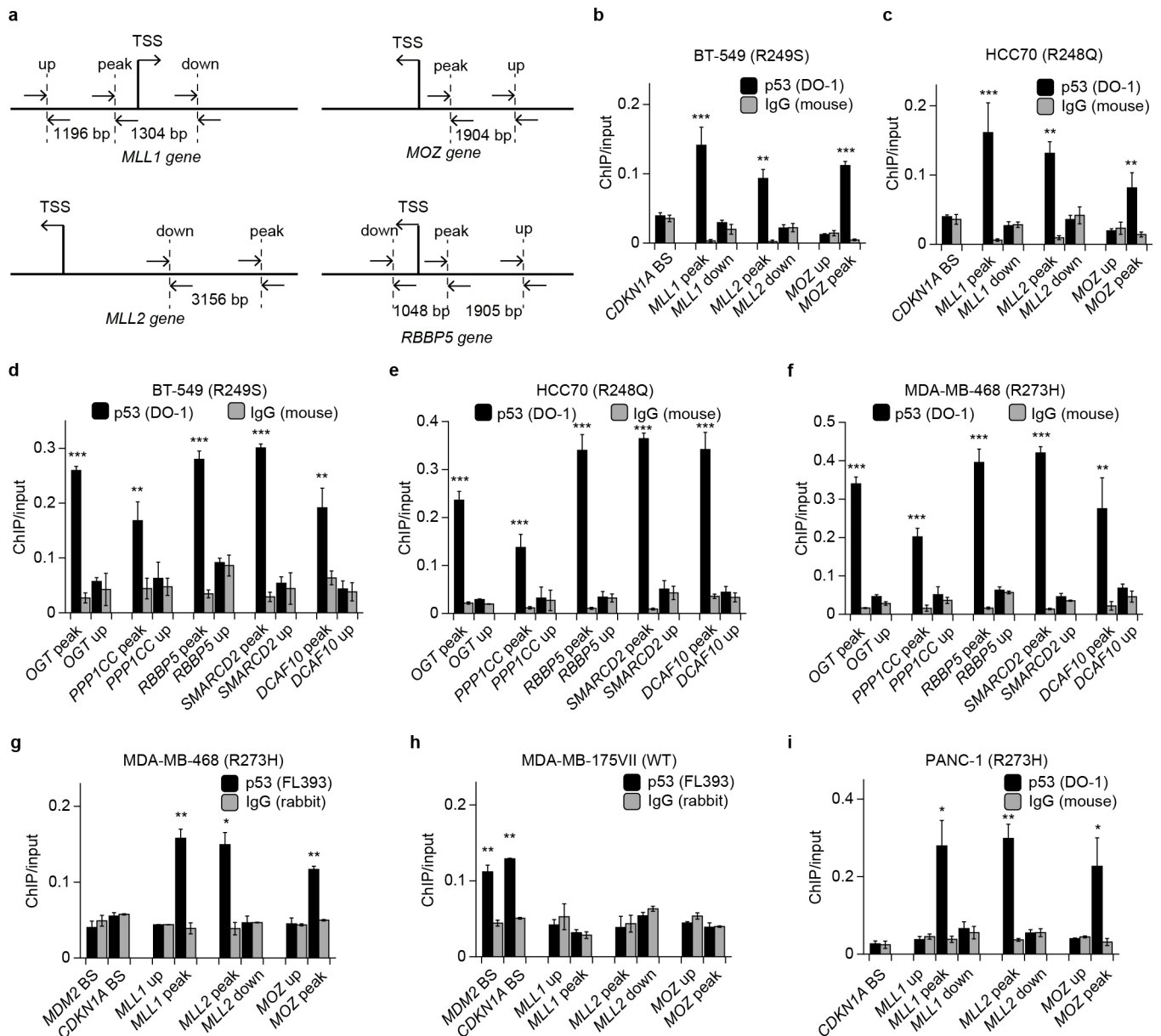
**Extended Data Figure 2 | GOF p53 genome-wide binding is in association with ETS family proteins.** **a**, Canonical ETS binding motif (top), and discovered motif from all TSS-proximal peaks in MDA-MB-468 predicted by MEME/TomTom (middle), or SeqPos (bottom). **b**, MEME/TomTom identified wild-type p53 motif from MDA-MB-175VII TSS-proximal peaks. **c**, GST pulldown of bacterially expressed GST or GST-ETS2 with *in vitro* translated wild-type p53 or p53(R175H). **d**, **e**, Co-immunoprecipitation at endogenous protein levels of ETS2 and GOF p53(R273H) (**d**) or wild-type p53

(**e**) in MDA-MB-468 (**d**) or MCF7 (**e**) cells. **f**, **g**, Box plots showing overlap of GOF p53 (**f**) TSS-proximal peaks from MDA-MB-468 cells or wild-type p53 (**g**) TSS-proximal peaks from MCF7 cells, with ETS family proteins (blue), all other transcription factors (grey) or Pol II (white) peaks from ENCODE ChIP-seq data sets. Whiskers on the box plots represent the inter-quartile range. Mann-Whitney *U*-tests were performed to compute significance. **h**, GO analysis of wild-type p53 TSS-proximal peaks (statistics are shown in Supplementary Table 1).



**Extended Data Figure 3 | UCSC Genome Browser views showing distinct wild-type p53 and GOF p53 binding patterns over representative canonical wild-type p53 targets and novel GOF p53 targets. a–d,** UCSC Genome Browser views of p53 occupancy over promoter regions of *MLL1* (a), *MLL2* (b), *CDKN1A* (c) and *MOZ* (d) in MCF7, HCC70, and BT-549 cells.

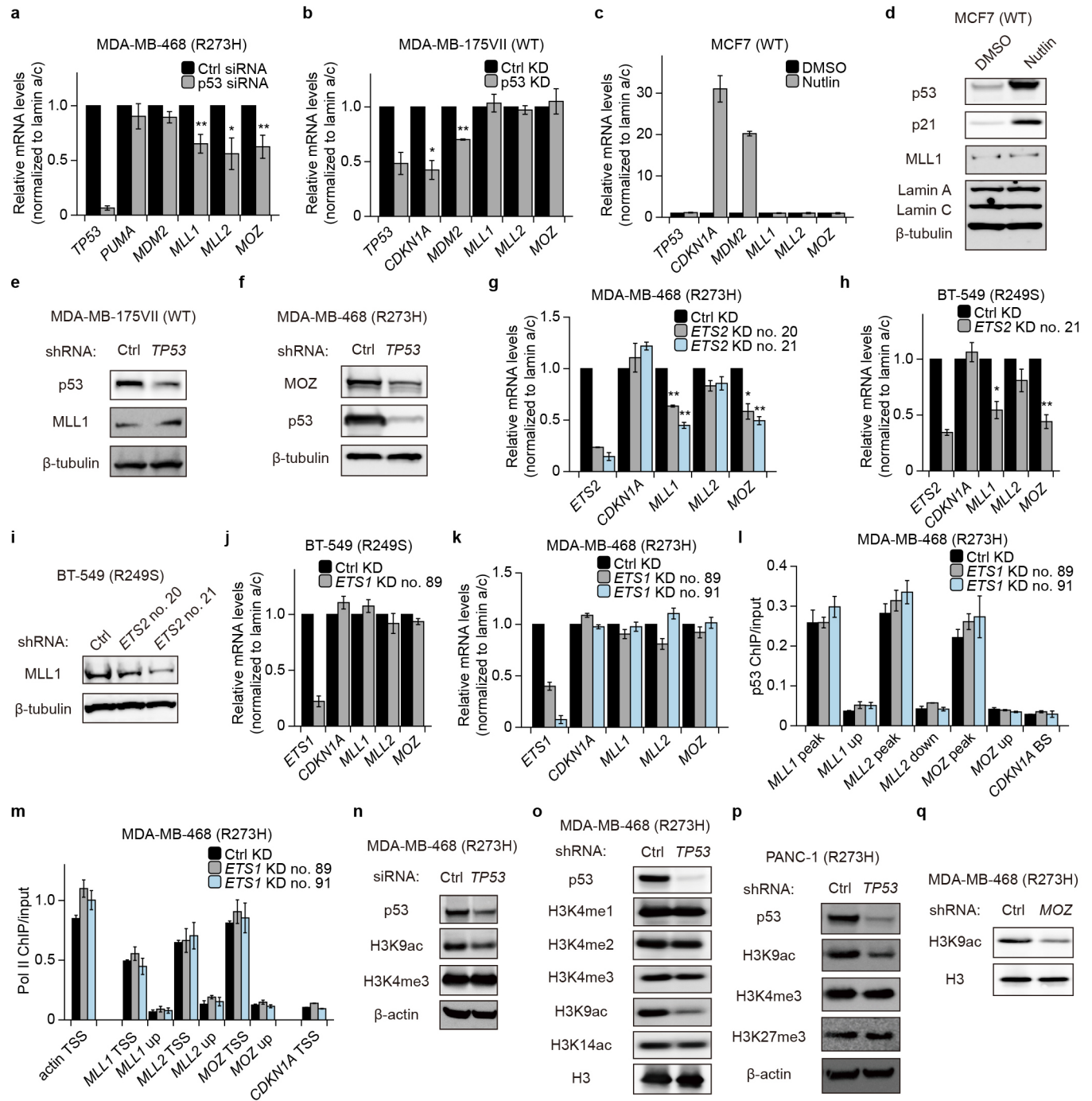
**e–g,** Re-aligned GOF p53(R248W) and IgG ChIP-seq data from LFS MDAH087 cells, showing enrichment of GOF p53 at promoter regions of *MLL1* (e), *MLL2* (f), and *MOZ* (g). **h–j,** UCSC Genome Browser views of p53 occupancy over promoter regions of *RBBP5* (h), *MDM2* (i) and *PUMA* (j), in MCF7, MDA-MB-175VII, HCC70, BT-549, and MDA-MB-468 cells.



**Extended Data Figure 4 | ChIP-qPCR validation of GOF p53 binding at newly identified chromatin regulator genes.** **a**, Schematic of amplicon locations for ChIP-qPCR validations performed in this study. **b**, **c**, ChIP-qPCR showing p53 (DO-1 antibody) or IgG (mouse) enrichment (ChIP/input) over *MLL1*, *MLL2* and *MOZ* peak regions, in BT-549 (**b**) and HCC70 (**c**) cells. **d**–**f**, ChIP-qPCR showing p53 (DO-1 antibody) or IgG (mouse) enrichment over *OGT*, *PPP1CC*, *RBBP5*, *SMARCD2*, and *DCAF10* peak regions in BT-549

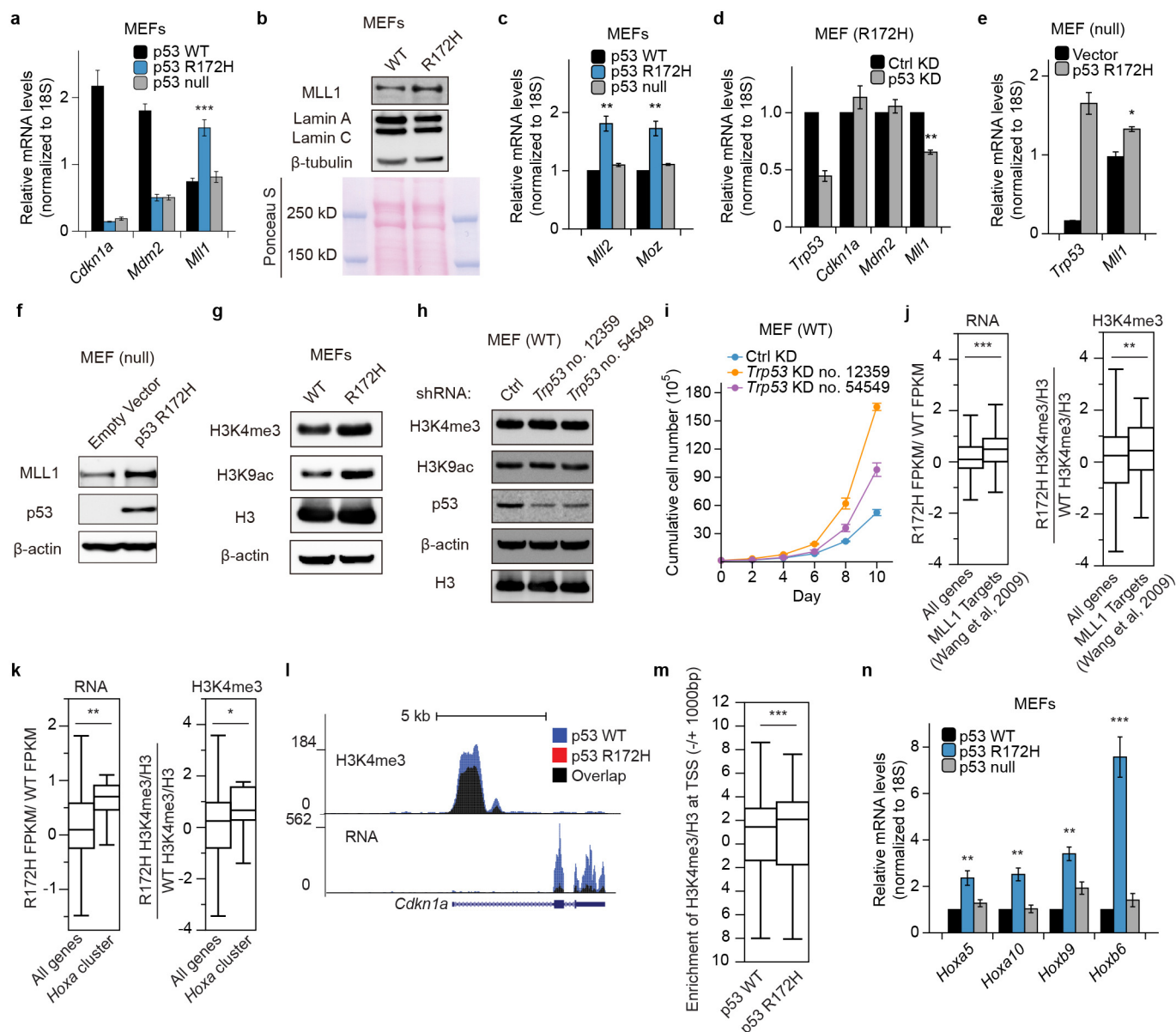
(**d**), HCC70 (**e**) and MDA-MB-468 (**f**) cells. **g**, **h**, ChIP-qPCR showing p53 (FL393 antibody) or IgG (rabbit) enrichment over *MDM2*, *CDKN1A*, *MLL1*, *MLL2* and *MOZ* regions, in MDA-MB-468 (**g**) and MDA-MB-175VII (**h**) cells. **i**, ChIP-qPCR showing p53 (DO-1 antibody) or IgG (mouse) enrichment over *MLL1*, *MLL2* and *MOZ* peak regions in PANC-1 cells. Error bars represent mean  $\pm$  s.e.m.;  $n = 3$ ; two-tailed Student's  $t$ -test: \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .





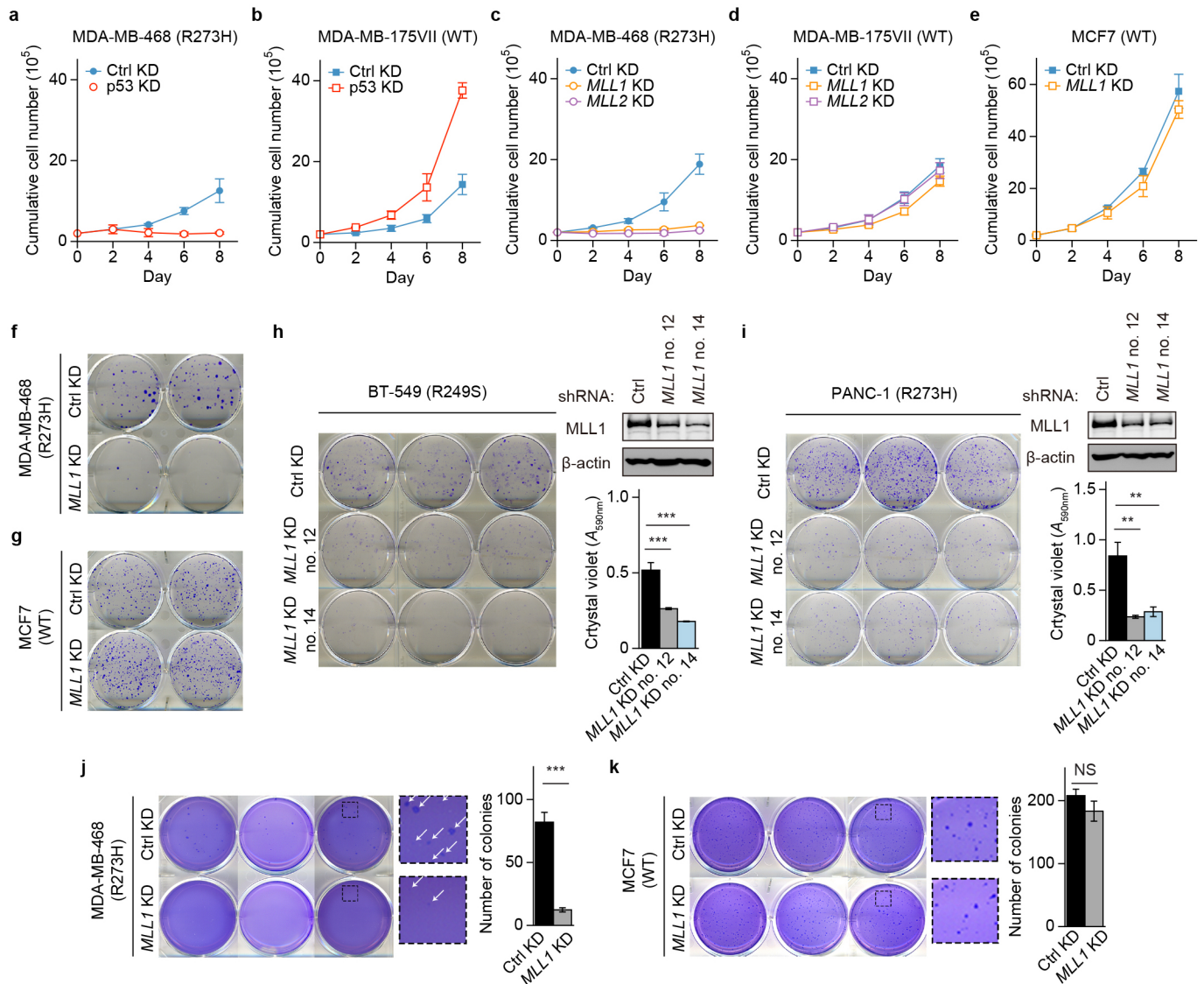
**Extended Data Figure 5 | GOF p53 regulates expression of MLL1, MLL2, and MOZ, and corresponding histone post-translational modifications in cancer cells.** **a, b**, RT-qPCR analysis measuring mRNA level changes upon siRNA-mediated GOF p53 knockdown in MDA-MB-468 cells (**a**), and shRNA-mediated wild-type p53 knockdown in MDA-MB-175VII cells (**b**). **c, d**, RT-qPCR analysis of mRNA levels (**c**), and western blot analysis of protein levels upon DMSO or nutlin treatment in MCF7 cells (**d**). **e**, Western blot analysis of MLL1 protein level upon shRNA-mediated wild-type p53 knockdown in MDA-MB-175VII cells. **f**, Western blot analysis of MOZ protein level change upon shRNA-mediated GOF p53 knockdown in MDA-MB-468 cells. **g**, RT-qPCR measuring mRNA levels changes upon shRNA-mediated ETS2 knockdown in MDA-MB-468 cells. **h, i**, RT-qPCR measuring mRNA levels (**h**) and western blot measuring protein levels (**i**) upon shRNA-mediated ETS2

knockdown in BT-549 cells. **j, k**, RT-qPCR measuring mRNA levels changes upon shRNA-mediated ETS1 knockdown in BT-549 (**j**) and MDA-MB-468 (**k**) cells. Numbers 89 and 91 denote two short hairpins targeting ETS1, sequences of which are shown in Supplementary Table 3. **l, m**, ChIP-qPCR showing p53 occupancy (**l**) and Pol II occupancy (**m**) upon shRNA-mediated ETS1 knockdown in MDA-MB-468 cells. **n, o**, Western blot analysis of histone methylation and acetylation level changes upon siRNA-mediated (**n**) or shRNA-mediated (**o**) knockdown of GOF p53 in MDA-MB-468 cells. **p**, Western blot analysis of histone methylation and acetylation level changes upon GOF p53 knockdown in PANC-1 cells. **q**, Western blot of H3K9ac change upon MOZ knockdown in MDA-MB-468 cells. Uncropped blots are shown in Supplementary Fig. 1. Error bars represent mean  $\pm$  s.e.m.; *n* = 3; two-tailed Student's *t*-test; \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.



**Extended Data Figure 6 | GOF p53 regulates expression of Mll1, Mll2, and Moz, and corresponding histone post-translational modifications in primary MEFs.** **a**, RT-qPCR analysis comparing *Mll1* expression levels between MEFs bearing wild-type p53, GOF p53(R172H), and p53 null. **b**, Western blot comparing Mll1 protein level between MEFs with wild-type p53 and GOF p53. **c**, RT-qPCR analysis comparing *Mll2* and *Moz* expression levels between MEFs bearing wild-type p53, GOF p53(R172H), and p53 null. **d**, RT-qPCR measuring mRNA changes upon shRNA-mediated p53 knockdown in GOF p53(R172H) knock-in MEFs. **e**, f, RT-qPCR analysis of mRNA levels (e) and western blot of protein levels (f) upon retroviral expression of GOF p53(R172H) in MEFs with p53 knockout. **g**, Western blot comparing H3K4me3 and H3K9ac levels between MEFs with wild-type p53 and GOF p53(R172H). **h**, Western blot showing H3K4me3 and H3K9ac levels upon p53 knockdown in wild-type p53 MEFs. **i**, Growth curve analysis of wild-type p53 MEF proliferation upon shRNA-mediated p53 knockdown. **j**, **k**, Box plot analysis of RNA levels (left) and H3 normalized H3K4me3 levels

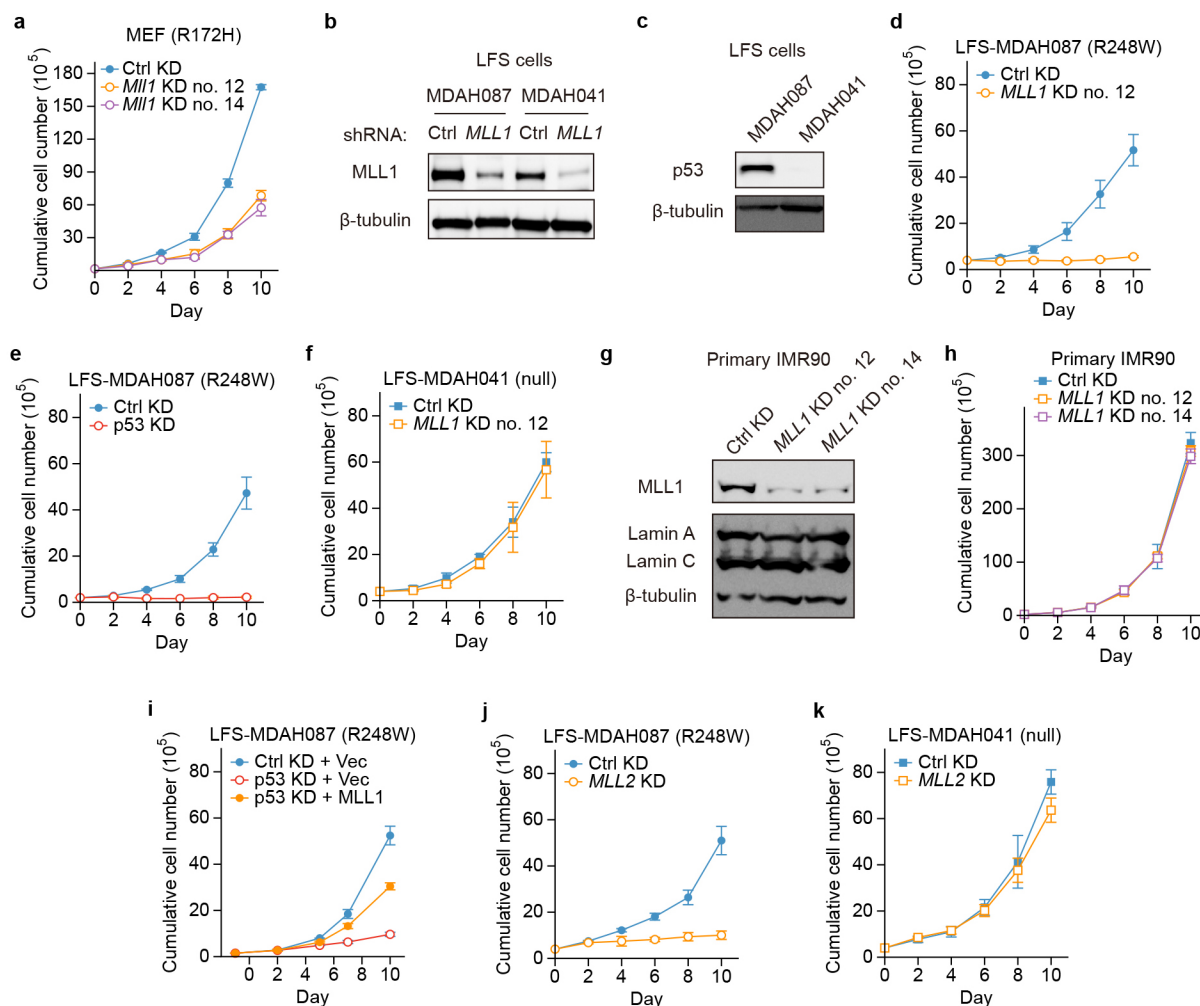
(right) at previously discovered Mll1 target genes (j) or *Hoxa* cluster genes (k) compared with all genes, from RNA-seq and H3K4me3 ChIP-seq in MEFs with wild-type p53 or GOF p53(R172H). Plots are presented as ratios of GOF p53(R172H) values over wild-type p53 values. **l**, UCSC Genome Browser views of H3K4me3 enrichment (top) and RNA levels (bottom) of *Cdkn1a*, from H3K4me3 ChIP-seq and RNA-seq of MEFs with wild-type p53 or GOF p53(R172H). Tracks are presented as overlay of wild-type p53 and GOF p53 signals. Blue denotes more enriched in wild-type p53, red denotes more enriched in GOF p53(R172H), black denotes overlap. **m**, Box plot of H3 normalized H3K4me3 levels over all gene TSSs, from H3K4me3 ChIP-seq in MEFs with wild-type p53 or GOF p53(R172H). **n**, RT-qPCR analysis comparing *Hox* gene expression levels between MEFs bearing wild-type p53, GOF p53(R172H), and p53 null. Uncropped blots are shown in Supplementary Fig. 1. For all bar graphs, two-tailed Student's *t*-test; \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001. Error bars represent mean  $\pm$  s.e.m.; *n* = 3. For all box plots, Mann-Whitney *U*-test; \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.



**Extended Data Figure 7 | MLL knockdown reduces proliferation and cancer phenotype specifically in GOF p53 cancer cells.** **a, b**, Growth curve analysis of MDA-MB-468 (**a**) and MDA-MB-175VII (**b**) cells with either non-targeting control shRNA or p53 shRNA knockdown. **c, d**, Growth curve analysis of MDA-MB-468 (**c**) and MDA-MB-175VII (**d**) cells with non-targeting control shRNA, *MLL1* shRNA, or *MLL2* shRNA knockdown. **e**, Growth curve analysis of MCF7 cells with non-targeting control shRNA or *MLL1* shRNA knockdown. **f, g**, Colony-formation assay of MDA-MB-468 (**f**) and MCF7 (**g**) cells with either non-targeting control shRNA or *MLL1* shRNA knockdown. Corresponding to Fig. 4a, **b, h, i**, Colony-formation assay of

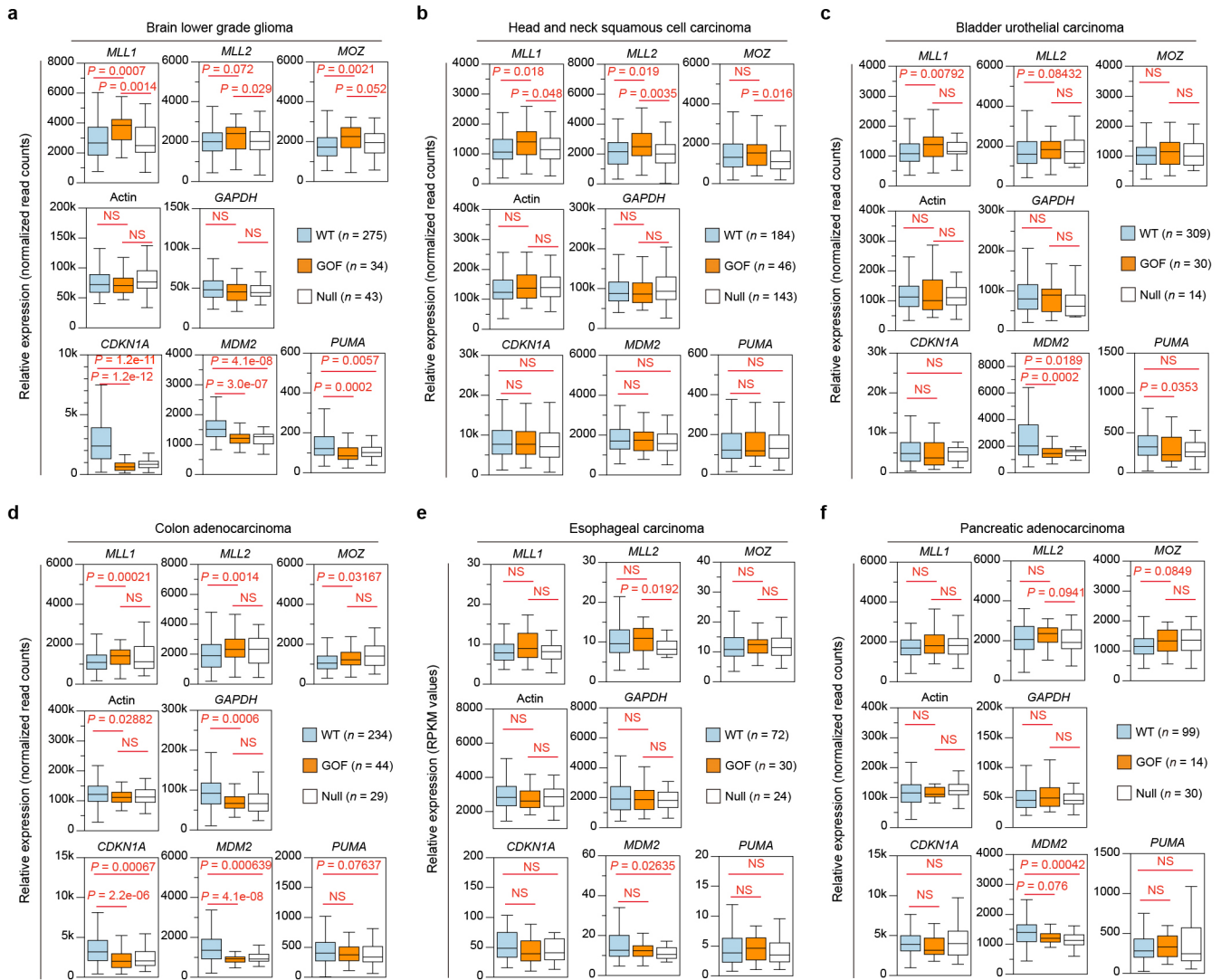
BT-549 (**h**) and PANC-1 (**i**) cells with either non-targeting control shRNA or two different *MLL1* shRNA knockdown, and quantification by crystal violet staining over three biological replicates. Reduction of *MLL1* protein is also shown by western blot. **j, k**, Anchorage-independent soft agar assay of MDA-MB-468 (**j**) and MCF7 (**k**) cells with either non-targeting control shRNA or *MLL1* shRNA knockdown. Dashed boxes denote enlarged images of the selected areas. White arrows indicate visible colonies in **j**. Quantifications are shown as number of visible colonies. Error bars represent mean  $\pm$  s.e.m.;  $n = 3$ ; two-tailed Student's *t*-test; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .





**Extended Data Figure 8 | MLL knockdown reduces proliferation specifically of GOF p53 MEFs and LFS cells.** **a**, Growth curve analysis of GOF p53(R172H) MEFs with either non-targeting control shRNA or two different *MLL1* shRNA knockdowns. **b**, Western blot analysis of MLL1 levels upon shRNA-mediated knockdown in LFS MDAH087 and MDAH041 cells. **c**, Western blot analysis of p53 protein levels in LFS MDAH087 and MDAH041 cells. **d**, **e**, Growth curve analysis of LFS MDAH087 cells upon *MLL1* (d) knockdown or p53 (e) knockdown. **f**, Growth curve analysis of LFS

MDAH041 cells upon *MLL1* knockdown. **g**, **h**, Western blot analysis of MLL1 level (g) and growth curve analysis (h) of proliferation upon shRNA-mediated *MLL1* knockdown in IMR90 cells. **i**, Growth curve analysis of LFS MDAH087 cells with non-targeting control shRNA plus empty vector, p53 shRNA plus vector, and p53 shRNA plus *MLL1* expressing vector. **j**, **k**, Growth curve analysis of LFS MDAH087 (j) and LFS MDAH041 (k) cells with either non-targeting control shRNA or *MLL2* shRNA knockdown.



**Extended Data Figure 9 | TCGA RNA expression profile analysis.**

**a–f.** TCGA RNA expression profile of GOF p53 target genes (top), housekeeping genes (middle), and wild-type p53 target genes (bottom) in brain lower grade glioma (**a**), head and neck squamous cell carcinoma (**b**), bladder urothelial carcinoma (**c**), colon adenocarcinoma (**d**), oesophageal

carcinoma (**e**) or pancreatic adenocarcinoma tumours (**f**) with wild-type p53 (blue), GOF p53 (orange), or p53 null (white). Expression values are normalized read counts (**a–d, f**), or RPKM values (**e**) from TCGA RNA-seq data sets. Mann–Whitney  $U$ -tests were performed to compute significance.

# An atomic structure of human $\gamma$ -secretase

Xiao-chen Bai<sup>1\*</sup>, Chuangye Yan<sup>2\*</sup>, Guanghui Yang<sup>2\*</sup>, Peilong Lu<sup>2</sup>, Dan Ma<sup>2</sup>, Linfeng Sun<sup>2</sup>, Rui Zhou<sup>2</sup>, Sjors H. W. Scheres<sup>1</sup> & Yigong Shi<sup>2</sup>

**Dysfunction of the intramembrane protease  $\gamma$ -secretase is thought to cause Alzheimer's disease, with most mutations derived from Alzheimer's disease mapping to the catalytic subunit presenilin 1 (PS1). Here we report an atomic structure of human  $\gamma$ -secretase at 3.4 Å resolution, determined by single-particle cryo-electron microscopy. Mutations derived from Alzheimer's disease affect residues at two hotspots in PS1, each located at the centre of a distinct four transmembrane segment (TM) bundle. TM2 and, to a lesser extent, TM6 exhibit considerable flexibility, yielding a plastic active site and adaptable surrounding elements. The active site of PS1 is accessible from the convex side of the TM horseshoe, suggesting considerable conformational changes in nicastrin extracellular domain after substrate recruitment. Component protein APH-1 serves as a scaffold, anchoring the lone transmembrane helix from nicastrin and supporting the flexible conformation of PS1. Ordered phospholipids stabilize the complex inside the membrane. Our structure serves as a molecular basis for mechanistic understanding of  $\gamma$ -secretase function.**

A hallmark of Alzheimer's disease is accumulation of  $\beta$ -amyloid plaque in the brains of patients<sup>1</sup>. The intramembrane protease  $\gamma$ -secretase is thought to contribute to the development of Alzheimer's disease by generating  $\beta$ -amyloid peptides (A $\beta$ s), particularly those that are prone to aggregation such as A $\beta$ 42 (refs 2–5). A mature  $\gamma$ -secretase consists of four components: presenilin, PEN-2, nicastrin, and APH-1 (ref. 6). Among these components, presenilin is responsible for the A $\beta$ -producing proteolytic activity<sup>7,8</sup>.

Presenilin comprises nine TMs, with the signature motifs YD on TM6 and GxGD on TM7 (ref. 8). During assembly of  $\gamma$ -secretase, presenilin undergoes an autocatalytic cleavage to produce an amino (N)-terminal fragment (NTF, comprising TMs 1–6) and a carboxy (C)-terminal fragment (CTF, comprising TMs 7–9)<sup>9</sup>. Among the 300 or so mutations derived from patients with familial Alzheimer's disease (FAD), more than two-thirds are mapped to PS1, and about three dozen each are derived from PS2 and amyloid precursor protein (APP). Nicastrin contains a large extracellular domain (ECD) and a single TM; its ECD is heavily glycosylated and thought to recognize the N terminus of substrate protein<sup>10–13</sup>. PEN-2 directly binds PS1 and is required for its autocatalytic maturation and protease activity<sup>14,15</sup>. APH-1 contains seven TMs and is indispensable for  $\gamma$ -secretase assembly<sup>16,17</sup>.

Our previous cryo-electron microscopy (cryo-EM) structure of human  $\gamma$ -secretase at 4.5 Å resolution led to identification of 19 TMs and construction of a partial atomic model for the ECD<sup>18</sup>. Analysis of the crystal structures of the archaeal presenilin homologue PSH<sup>19</sup> and the nicastrin homologue from *Dictyostelium purpureum* (DpNCT)<sup>20</sup> yielded a tantalizing clue on TM assignment and an improved atomic model for the ECD. A subsequent cryo-EM structure of human  $\gamma$ -secretase at 4.3 Å resolution allowed assignment of all 20 TMs to the four components<sup>21</sup>. In this study, we report the first atomic structure of an intact human  $\gamma$ -secretase, which allows visualization of the atomic structures of all four components and determination of the specific interactions that underlie  $\gamma$ -secretase assembly.

## Structure determination of human $\gamma$ -secretase

We performed cryo-EM single-particle analysis on the same sample in amphipols that was used to calculate our previous map<sup>18</sup>. With the aim of reaching higher resolution, we used zero-loss energy-filtered imaging, a higher magnification, and a lower dose rate on the single-electron counting detector (see Methods). We also collected a larger data set. An initial set of 412,272 particle images yielded a 3.5 Å map. Subsequent three-dimensional (3D) classification that was focused on the TMs resulted in a more structurally homogeneous subset of 159,549 particles. This subset was used to calculate the final map at 3.4 Å resolution (Extended Data Figs 1 and 2). Our map displays excellent main-chain connectivity and side-chain densities for almost all residues (Extended Data Fig. 2a).

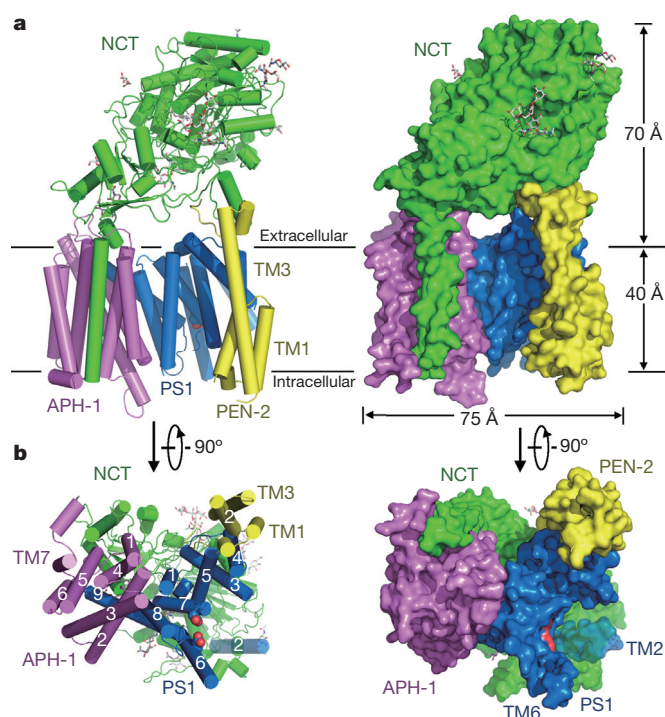
Twenty TMs were identified in the transmembrane region, including a highly mobile TM (TM2 of PS1) that is only visible as rod-shaped density in a 7 Å low-pass filtered map (Extended Data Fig. 2b). This low-pass filtered map also shows a second rod-shaped density in the PS1 cavity between TM3, TM5, and TM2, which we were unable to identify. As previously reported<sup>20,21</sup>, PS1 and APH-1 are located at the centre of the TM horseshoe. All seven TMs of APH-1 have exceptional density, with most side chains clearly identifiable (Extended Data Fig. 2c). Except for TM2 and TM6, all other TMs of PS1 exhibit excellent density, with aromatic and bulky residues easily recognizable (Extended Data Fig. 2d). The three TMs of PEN-2 on the thin end of the TM horseshoe and the lone TM of nicastrin on the thick end both display discernible side-chain features (Extended Data Fig. 2e, f). Several N-linked glycans and lipid molecules are also defined by clear EM density (Extended Data Fig. 2g, h).

On the basis of these unambiguous EM densities, we built and refined a near-complete atomic model for human  $\gamma$ -secretase (Fig. 1 and Extended Data Table 1), which includes 598 residues in the TMs and 632 residues in the ECD. Assignment of specific residues in the TMs was aided by a large number of aromatic amino acids (Extended Data Fig. 2c–f). The density map for PS1-TM2 is inadequate for

<sup>1</sup>MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. <sup>2</sup>Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.

\*These authors contributed equally to this work.





**Figure 1 | Atomic structure of human  $\gamma$ -secretase.** **a**, The  $\gamma$ -secretase structure is shown in cartoon representation (left) and surface view (right). Eleven N-linked glycans are displayed in stick. **b**, The  $\gamma$ -secretase structure is viewed perpendicular to the lipid membrane from the intracellular side. TM2 of PS1 is most flexible and shown in a semi-transparent fashion. The catalytic residues Asp257 and Asp385 are located on the convex side of the TM horseshoe. All structural figures were prepared using UCSF Chimera<sup>35</sup> or PyMol<sup>36</sup>.

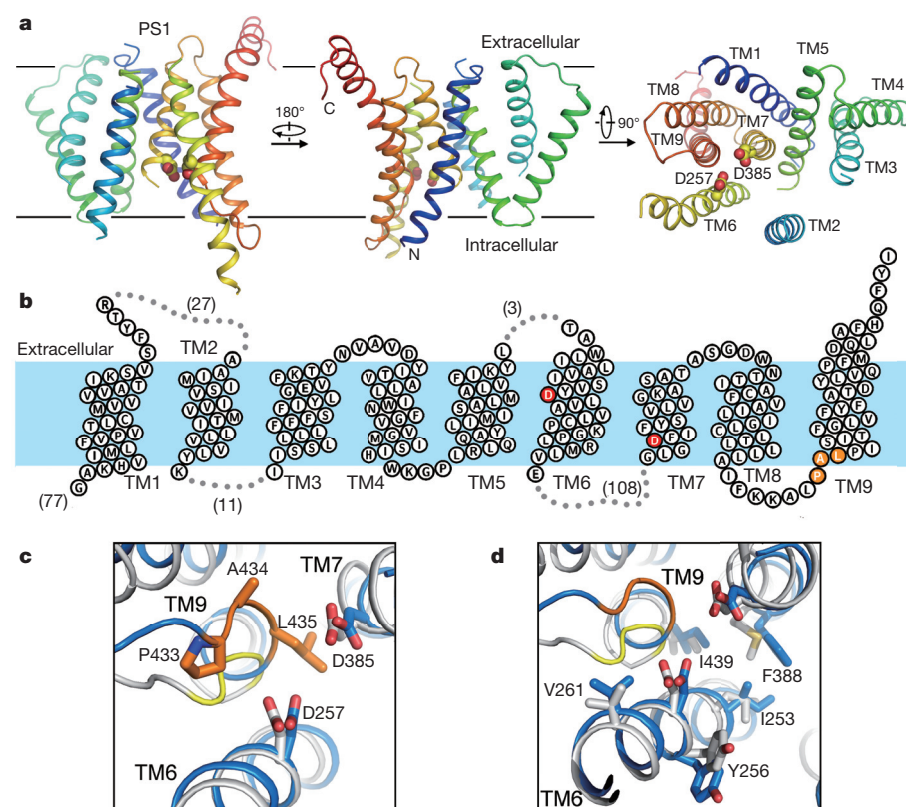
model building; nonetheless, we generated an atomic model for this TM on the basis of its sequence homology to the archaeal homologue PSH<sup>19</sup>. This is the first structure of human  $\gamma$ -secretase at a near-atomic resolution (Fig. 1 and Extended Data Fig. 3).

Nicastrin ECD, which constitutes the bulk of the extracellular region of human  $\gamma$ -secretase, contains 11 glycans and directly interacts with two ends of the TM horseshoe (Fig. 1a). The lone TM of nicastrin closely stacks against TM1/TM5/TM7 of APH-1 through predominantly hydrophobic interactions at the thick end. TM2 and TM4 of APH-1 associate with TM8 and TM9 of PS1 at the centre of the TM horseshoe, with the C-terminal three residues Phe465–Tyr466–Ile467 of PS1 inserting into a greasy central cavity on the extracellular side of APH-1 (Fig. 1a). The three TMs of PEN-2 bind to TM4 of PS1 at the thin end. Notably, this arrangement places the two catalytic residues, Asp257 and Asp385 of PS1, on the convex side of the TM horseshoe (Fig. 1b).

### Atomic structure of presenilin

PS1 exhibits an extended organization, with empty spaces between a few adjacent TMs (Fig. 2a and Extended Data Fig. 4). The N-terminal 77 residues have no EM density and are presumably disordered, probably because of their intrinsic flexibility. Among the seven surface loops that connect neighbouring TMs, four exhibit clear and contiguous density. The extended sequences between TM6 and TM7 harbour the site of autocatalytic cleavage; but these sequences are hydrophilic and mostly disordered (Fig. 2b). The nine TMs exhibit a large variation in length, with TM9 containing 30 residues and TM7 only 18.

Among the two catalytic residues, Asp257 is located in the middle of TM6 slightly to the extracellular side, whereas Asp385 maps to the cytoplasmic side of TM7 (Fig. 2b). The distance between the C $\alpha$  atoms of Asp257 and Asp385 measures approximately 10.6 Å (Fig. 2c), considerably longer than that in an activated aspartate protease such as pepsin<sup>22</sup>. Importantly, however, these catalytic residues are placed



**Figure 2 | Atomic structure of PS1.** **a**, PS1 has a loosely organized structure and exhibits considerable flexibility. The cartoon representation of PS1 is rainbow-coloured. TM2 is visible only at low resolutions, and the density map contains no features for side chains. Nonetheless, an atomic model for TM2 was built on the basis of sequence and structural homology between PS1 and PSH<sup>19,21</sup>. **b**, A membrane topology diagram of PS1. The two catalytic aspartate residues are coloured red. **c**, The two catalytic aspartate residues of PS1 are in near-perfect registry with those in PSH<sup>19</sup>. The PAL sequence motif implicated in substrate recognition is shown. **d**, PS1 and PSH share similar features at their active sites.

next to the PAL signature motif on TM9 that is thought to play a role in substrate recognition<sup>23–25</sup> (Fig. 2c). We speculate that substrate binding may trigger alignment of these two aspartate residues and consequent catalysis.

Despite a relatively low sequence identity of 19% between PS1 and the archaeal intramembrane protease PSH, their overall structures are similar to each other<sup>19,21</sup>. In particular, the catalytic residues are in nearly perfect registry between PS1 and PSH (Fig. 2c). The amino acids that surround the catalytic residues, including the PAL motif, are also highly conserved. Relative to Asp257, three residues of PS1 (Ile253, Tyr256, and Val261) are located on the same side of TM6; these residues are replaced by Leu158, Tyr161, and Val166, respectively, in PSH (Fig. 2d). Similarly, Gly382 and Phe388 of PS1 correspond to Gly219 and Met223 of PSH, respectively. Ile439 on TM9 of PS1 and Ile283 of PSH nearly coincide with each other (Fig. 2d). The observed structural conservation may underlie the finding that PSH exhibits similar cleavage preferences towards APP C99 as human  $\gamma$ -secretase<sup>26</sup>.

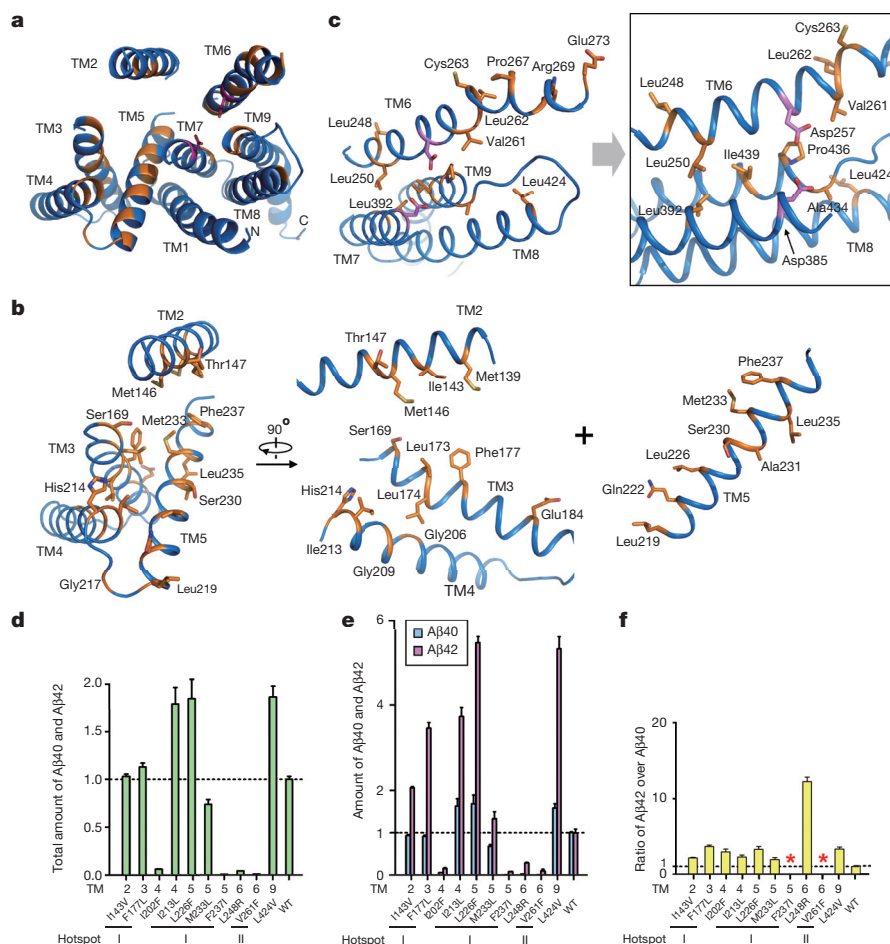
### Mutational hotspots in PS1 structure

The PS1 structure at a near-atomic resolution allows mapping and analysis of disease-derived mutations in PS1. On the basis of public information (<http://www.alzforum.org/mutations>), the 212

Alzheimer's-disease-derived missense mutations in PS1 affect 135 amino acids, of which 53 are targeted for two or more mutations. Presumably these residues affect the specific function of PS1 that is directly related to development of Alzheimer's disease. Among these 53 residues, 35 are identified in the TMs of our current PS1 structure, and the rest are located on surface elements that have no EM density. Notably, there are eight such residues in TM5, seven in TM6, but none in TM1. Together, these 35 amino acids account for a total of 101 Alzheimer's-disease-derived missense mutations.

Analysis of these 35 residues on the structure of PS1 led to identification of two mutational hotspots, each located at the inner core of a structural repeat (Fig. 3a). The first hotspot involves the inner core of TMs 2–5 (Fig. 3b). Among the 20 affected residues in these TMs, 18 have their side chains facing the inner core of TMs 2–5. Consequently, only one side of each TM helix is affected. For example, Leu219, Glu222, Leu226, Ser230, Met233, and Phe237 are placed on the same side of TM5 (Fig. 3b). Similarly, the second hotspot is located at the inner core of TMs 6–9, in the vicinity of the catalytic residues Asp257 and Asp385 (Fig. 3c).

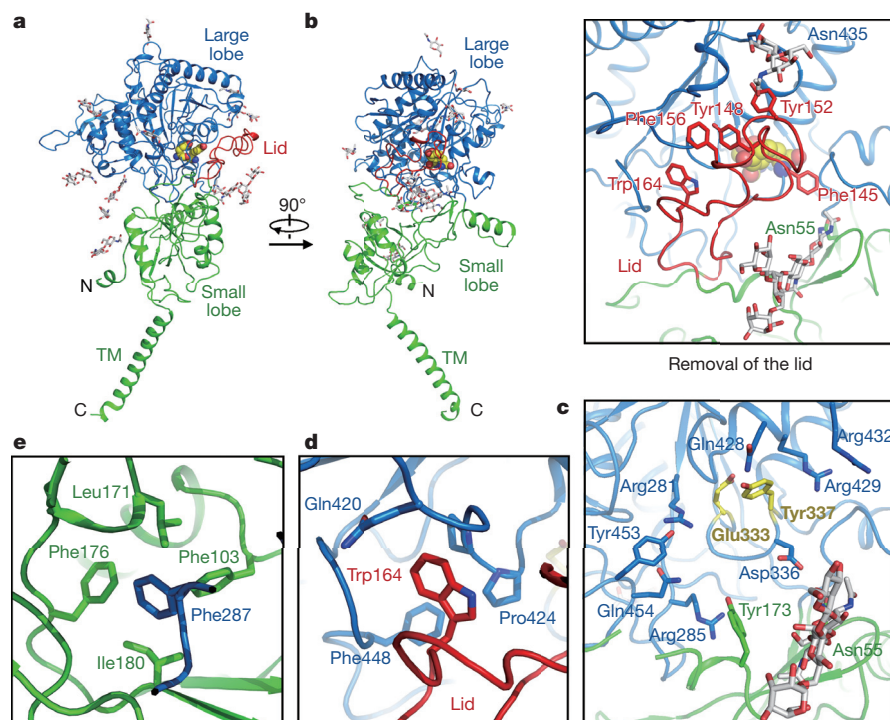
To examine the functional consequence of disease-derived mutations, we generated ten  $\gamma$ -secretase variants, each containing a distinct missense mutation in PS1. Among these mutations, seven map to the two mutational hotspots. These  $\gamma$ -secretase variants were individually



**Figure 3 | Alzheimer's disease-derived mutations map to two hotspots in PS1.** **a**, An overall view of the PS1 residues targeted for mutations in patients with Alzheimer's disease. PS1 is viewed from the extracellular side. Mutated residues are coloured orange. **b**, Close-up views of the mutation-targeted residues in TMs 2–5. Most of these residues map to the centre of this four-TM bundle. **c**, Close-up views of the mutation-targeted residues in TMs 6–9. The two catalytic residues Asp257 and Asp385 are shown. **d**, FAD-derived mutations in PS1 have varying degrees of effect on the combined Aβ40 and

Aβ42 cleavage activity of  $\gamma$ -secretase. Shown here are results of ten such  $\gamma$ -secretase mutants, each containing a specific mutation derived from FAD. The activity of WT  $\gamma$ -secretase is normalized to 1.0. **e**, FAD-derived mutations either suppressed the production of Aβ40 more than Aβ42 or increased the production of Aβ40 less than Aβ42. **f**, All but two FAD-derived mutations led to increased Aβ42/Aβ40 ratios. Two mutations F237I and V261F in PS1 abrogated Aβ40 cleavage altogether, disallowing calculation of the Aβ42/Aβ40 ratio. Each experiment was independently repeated three times. Error bars, s.d.





**Figure 4 | Structural features of nicastrin.** **a**, Two perpendicular views of nicastrin. The lid from the small lobe is highlighted in red. Surface glycans are shown. **b**, The lid hovers above a hydrophilic pocket in the large lobe. Two large glycans on Asn55 and Asn435 sandwich the lid and interact with surrounding residues. **c**, Glu333 and Tyr337 are surrounded by several charged and polar residues in the pocket. These structural features are consistent with the pocket being a binding site for substrate protein. **d**, Trp164 from the lid makes van der Waals contacts to Pro424, Phe448, and the aliphatic side chain of Gln420. **e**, Phe287 from the large lobe may serve as the hydrophobic pivot. Phe287 interacts with four hydrophobic residues from the small lobe.

purified to homogeneity and examined for their protease activities towards the APP C99 substrate in an *in vitro* cleavage assay. Compared with the wild-type (WT)  $\gamma$ -secretase, four mutations (I202F, F237I, L248R, and V261F) lead to severely compromised protease activity (Fig. 3d). Of these four mutations, F237I and V261F failed to generate any detectable amount of A $\beta$ 40 (Fig. 3e), disallowing calculation of the A $\beta$ 42/A $\beta$ 40 ratio (Fig. 3f). In contrast, three mutations (I143V, F177L, and M233L) had relatively little impact on the total cleavage activity of  $\gamma$ -secretase (Fig. 3d), whereas the remaining three mutations (I213L, L226F, and L424V) actually increased the total cleavage activity (Fig. 3d). These observations strongly suggest a disconnection between the total protease activity of  $\gamma$ -secretase and the development of FAD (where these mutations were isolated). All eight mutations for which the A $\beta$ 42/A $\beta$ 40 ratio can be calculated led to increased A $\beta$ 42/A $\beta$ 40 ratios compared with WT  $\gamma$ -secretase (Fig. 3f). The generally increased ratios of A $\beta$ 42 over A $\beta$ 40 may suggest a causal relationship for development of FAD, but could also be explained by other possibilities.

### Atomic structure of nicastrin

Similar to DpNCT<sup>20</sup>, human nicastrin also contains a large lobe, a small lobe, and a lone TM (Fig. 4a and Extended Data Fig. 5). These two nicastrin structures can be aligned to each other with a root mean squared deviation of 2.2 Å. Of the 230 kDa molecular mass in mature  $\gamma$ -secretase, up to 70 kDa can be attributed to glycosylation of nicastrin<sup>27</sup>. Among the 16 predicted N-linked glycosylation sites in nicastrin, at least 11 are glycosylated, as judged by the EM density. Because the extended glycans are flexible in solution, only a small portion close to each Asn residue was modelled (Fig. 4a).

Nicastrin is thought to rely on Glu333 and Tyr337 for substrate recruitment<sup>10–13</sup>. In the structure of human nicastrin, Glu333 and Tyr337 are buried in a hydrophilic pocket that is covered by an extended surface loop known as the lid<sup>20</sup> (Fig. 4b). The lid, containing five aromatic residues, is sandwiched by two prominent glycans on Asn55 and Asn435 (Fig. 4b). Several charged and polar amino acids are located in the pocket, including four arginine residues: Arg281, Arg285, Arg429, and Arg432 (Fig. 4c). With the potential to mediate specific interactions such as hydrogen bonds, buried charged and/or

polar residues often serve a functional role. These structural features support the notion that this pocket is responsible for substrate recruitment, with these residues directly contributing to recognition.

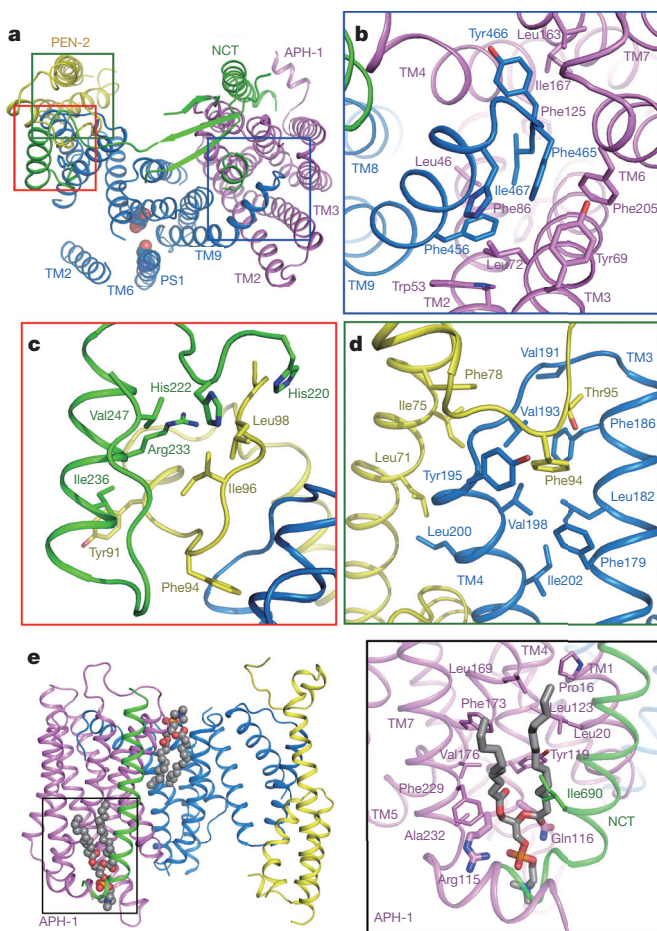
The closed conformation of the lid is sustained by specific interactions between residues from the lid and residues from surrounding structural elements. For example, the indole ring of Trp164 makes several van der Waals contacts to the side chains of Pro424 and Phe448 and the aliphatic portion of Gln420 (Fig. 4d). Substrate binding requires opening of the lid and disruption of these interactions. Because the lid comes from the small lobe, a rotation of the large lobe around a central pivot—Phe287—is proposed to be both necessary and sufficient for lid opening<sup>20</sup>. Such a rotation around Phe287 would be greatly facilitated by hydrophobic interactions, which are more adaptive to conformational changes than hydrogen bonds. Consistent with this analysis, Phe287 is nestled in a greasy pocket formed by Phe103, Leu171, Phe176, and Ile180 from the small lobe (Fig. 4e).

### Inter-component interactions

Assembly of the four components into a functional human  $\gamma$ -secretase involves specific interactions mainly in the membrane-spanning region, resulting in the burial of approximately 3,794 Å<sup>2</sup> of otherwise membrane-exposed surface areas (Fig. 5a). Binding of the ECD onto the TMs of  $\gamma$ -secretase involves an additional buried surface area of 1,320 Å<sup>2</sup>. The intra-membrane interactions comprise predominantly van der Waals contacts among hydrophobic amino acids, exemplified by the insertion of three residues at the C terminus of PS1 into a greasy pocket formed by TM2/3/4/6/7 of APh-1 (Fig. 5b). Phe465 from PS1 contacts Leu72 and Phe205 in APh-1, whereas Tyr466 from PS1 stacks against Leu163 and Ile167 in APh-1. The aliphatic side chain of Ile467 from PS1 interacts with Leu46, Leu86, and Phe125 at the bottom of the cavity in APh-1. The C terminus of PS1 is in close proximity to the highly conserved His171 and His197 in APh-1, which were found to be important for APh-1 binding to  $\gamma$ -secretase<sup>28</sup>. Interestingly, a peptide comprising the C-terminal eight residues of PS1 was shown to inhibit A $\beta$  production<sup>29</sup>, presumably by weakening or disrupting normal assembly of  $\gamma$ -secretase.

On the thin end of the TM horseshoe, an  $\alpha$ -helix and its preceding and ensuing loops in nicastrin stack against an extended loop at the





**Figure 5 | Assembly interfaces among the four components of  $\gamma$ -secretase in the transmembrane region.** **a**, An overall view of the packing interfaces in the transmembrane region. The three boxed interfaces are detailed in **b–d**. **b**, The C-terminal three residues Phe465–Tyr466–Ile467 of PS1 insert into a cavity formed by TMs in APH-1. **c**, Nicastrin interacts with PEN-2 through van der Waals contacts on a flat interface. **d**, PEN-2 binds to PS1 mostly through van der Waals contacts. In particular, Phe94 of PEN-2 is nestled in the greasy pocket of PS1, formed by Phe179, Leu182, Phe186, Val193, Tyr195, and Val198. **e**, Two phospholipids appear to stabilize the inter-component interfaces in  $\gamma$ -secretase. One lipid is bound at the interface between PS1 and APH-1, whereas the other is intercalated between the lone TM of nicastrin and TMs 1/4/5/7 of APH-1 (left). The aliphatic tails of the latter phospholipid may interact with several hydrophobic residues whereas its phosphate group likely hydrogen bonds to Arg115 and Gln116 (right).

C-terminal end of PEN-2 through multiple van der Waals contacts (Fig. 5c). At the nearby interface between PS1 and PEN-2, Phe94 from PEN-2 is nestled in a hydrophobic pocket formed by six residues in PS1 (Fig. 5d). Hydrophobic residues from TM3 of PEN-2, exemplified by Leu71, Ile75, and Phe78, also interact with residues in PS1. Notably, although PEN-2 has only three TMs and interacts with both PS1 and nicastrin, it actually contains three hydrophobic structural cores (Extended Data Fig. 6).

At 3.4 Å resolution, at least two phospholipid molecules were identified, each with two aliphatic tails linked to a small head group (Extended Data Fig. 2h). One lipid binds to the interface between PS1 and APH-1, making close contacts with residues in TM1/TM8 of PS1 and TM4 of APH-1 (Fig. 5e). The other lipid molecule is intercalated into the interface between APH-1 and the lone TM of nicastrin (Fig. 5e, right panel). The aliphatic tails of this phospholipid interact with several surrounding hydrophobic residues, whereas the phosphate group may directly hydrogen bond to the side chains of Arg115 and Gln116 from APH-1.

This lipid probably stabilizes the nicastrin–APH-1 interface within the membrane.

To corroborate the observed inter-component interfaces, we generated eight mutant  $\gamma$ -secretase complexes, each containing two specific cysteine mutations on two neighbouring components to facilitate formation of designed disulfide bonds. Three such  $\gamma$ -secretase mutants target the interface between APH-1 and nicastrin (Extended Data Fig. 7a). The distance between the C $\alpha$  atoms of Val147 on APH-1 and Ile40 on nicastrin is 4.1 Å, ideal for formation of a disulfide bond if these two residues were replaced by cysteine. Indeed, in the absence of the reducing agent DTT, APH-1 V147C was crosslinked to nicastrin I40C within the mutant  $\gamma$ -secretase, resulting in a high molecular mass complex on denaturing SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (Extended Data Fig. 7a, lane 1). This crosslinked complex disappeared in the presence of DTT (lane 2) and was absent for the WT  $\gamma$ -secretase regardless of DTT (lanes 7 and 8). Similarly, two additional  $\gamma$ -secretase mutants, one containing APH-1 V146C and nicastrin A664C and the other containing APH-1 A4C and nicastrin L673C, allowed specific crosslinking only in the absence of DTT (lanes 3–6). The same strategy was successfully applied to verify the specific interactions at the interfaces between PS1 and PEN-2 (Extended Data Fig. 7b), between PEN-2 and nicastrin (Extended Data Fig. 7c), and between APH-1 and PS1 (Extended Data Fig. 7d).

## Discussion

In this paper, we report the cryo-EM structure of human  $\gamma$ -secretase at an overall resolution of 3.4 Å. The qualitative improvement in resolution over earlier studies allowed us to derive an atomic model for all four components of  $\gamma$ -secretase. Except nicastrin<sup>18,20</sup>, such atomic models are reported here for the first time.

The available structural evidence supports the notion that Glu333 and Tyr337 of nicastrin, both buried in a hydrophilic pocket, may play a key role in substrate recruitment<sup>12</sup>. Under this scenario, displacement of the lid, caused by a movement of the large lobe relative to the small lobe, is required before substrate binding<sup>20</sup>. This proposal is supported by the unique pattern of contacts between the large and small lobes and by the conversed interactions around the central pivot Phe287 in nicastrin (Fig. 4). In substrate-free  $\gamma$ -secretase, the lid is positioned right above the concave side of the TM horseshoe whereas the active site is located on the convex side, resulting in a relatively long distance between the putative binding pocket for the N terminus of the substrate and the site of cleavage by PS1 (Extended Data Fig. 8). This distance may shorten, however, in response to conformational changes within  $\gamma$ -secretase. In particular, rotation of the large lobe relative to the small lobe, induced by substrate binding, may also re-orient the substrate for cleavage, perhaps by aligning the pocket in nicastrin and the active site in PS1. Dynamic conformations of  $\gamma$ -secretase have been observed<sup>30</sup>, particularly in the ECD region<sup>31</sup>.

TM6 of PS1, which harbours the catalytic residue Asp257, exhibits relatively poor EM density, whereas TM2 is visible only at low resolutions. These structural observations are consistent with the notion that binding by inhibitors or modulators may induce pronounced conformational changes<sup>30–32</sup>. In particular, substrate binding may trigger a conformational change, rendering the active site suitable for catalysis<sup>32</sup>. The plasticity observed in our structure may expedite the conformational changes that are needed to bring the two catalytic aspartates within hydrogen bonding distance of each other, and may play a role in the relaxed substrate specificity of the complex. Activation of the active site also depends on the binding of PEN-2, which was observed to have an allosteric effect on TM6 (ref. 33).

The prevailing  $\beta$ -amyloid hypothesis suggests an increased ratio of A $\beta$ 42 over A $\beta$ 40 as the major culprit for the accumulation of  $\beta$ -amyloid plaque and consequent development of Alzheimer's disease<sup>2,3</sup>. Our finding that FAD-derived mutations have varying degrees of effect on the protease activity of  $\gamma$ -secretase argues against a causal

relationship between the total protease activity of  $\gamma$ -secretase and the development of FAD. Our finding is consistent with the poor clinical performance of  $\gamma$ -secretase-inhibiting drugs<sup>34</sup>. Notably, the generally increased ratios of A $\beta$ 42 over A $\beta$ 40 only suggest, but fail to prove, a direct causal relationship for the development of FAD. For example, the assumption that  $\gamma$ -secretase has evolved to optimize production of A $\beta$ 40 over all other peptides would naturally lead to the prediction that any mutation, disease-causing or not, will increase the A $\beta$ 42/A $\beta$ 40 ratio.

Intriguingly, the mutations causing Alzheimer's disease map to two hotspots, each located at the inner core of a four-TM bundle (Fig. 3). The mutations on TMs 6–9 affect residues in the immediate neighbourhood of the catalytic residues Asp257 and Asp385 and thus may cripple the protease activity of  $\gamma$ -secretase. The biochemical data seem to support this analysis (Fig. 3d). However, the mutations on TMs 2–5 defy such rationale: they either abrogate or increase the protease activity of  $\gamma$ -secretase. The generally inward-facing nature of the mutation-targeted residues on TMs 2–5 (Fig. 3b) may suggest a transport function of some sort, or a binding site for another transmembrane protein. These tantalizing clues await experimental examination.

Our structure represents a milestone for the rapidly emerging technique of high-resolution cryo-EM structure determination. Because the signal-to-noise ratio in cryo-EM images correlates with the size of the particles, determining the relative orientations of individual particles becomes a limitation for small complexes. With a protein mass of ~170 kDa,  $\gamma$ -secretase is much smaller than any other atomic-resolution cryo-EM structure determined so far, while its lack of symmetry further complicates structure determination. Whereas glycosylation adds another 30–70 kDa of mass to the complex, only a small proportion of the sugar moieties are ordered and their effect on alignment is limited. Although glycosylation may inhibit crystal growth, it did not seem to hamper cryo-EM structure determination. On the contrary, glycosylation may stabilize the protein, possibly by protecting the protein from the denaturing air–water interface of the thin cryo-EM sample.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 May; accepted 6 July 2015.**

**Published online 17 August 2015.**

- Alzheimer, A. About a peculiar disease of the cerebral cortex. *Zbl. Nervenheilk. Psychiatrie* **30**, 177–179 (1907).
- Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184–185 (1992).
- De Strooper, B., Iwatsubo, T. & Wolfe, M. S. Presenilins and  $\gamma$ -secretase: structure, function, and role in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006304 (2012).
- Goate, A. & Hardy, J. Twenty years of Alzheimer's disease-causing mutations. *J. Neurochem.* **120** (suppl. 1), 3–8 (2012).
- Tanzi, R. E. & Bertram, L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* **120**, 545–555 (2005).
- Kimberly, W. T. et al.  $\gamma$ -Secretase is a membrane protein complex comprised of presenilin, nicastrin, aph-1, and pen-2. *Proc. Natl Acad. Sci. USA* **100**, 6382–6387 (2003).
- De Strooper, B. et al. Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**, 387–390 (1998).
- Wolfe, M. S. et al. Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and  $\gamma$ -secretase activity. *Nature* **398**, 513–517 (1999).
- Thinakaran, G. et al. Endoproteolysis of presenilin 1 and accumulation of processed derivatives in vivo. *Neuron* **17**, 181–190 (1996).
- Yu, G. et al. Nicastrin modulates presenilin-mediated notch/glp-1 signal transduction and  $\beta$ APP processing. *Nature* **407**, 48–54 (2000).
- Shah, S. et al. Nicastrin functions as a  $\gamma$ -secretase-substrate receptor. *Cell* **122**, 435–447 (2005).
- Dries, D. R. et al. Glu-333 of nicastrin directly participates in  $\gamma$ -secretase activity. *J. Biol. Chem.* **284**, 29714–29724 (2009).
- Goo, J. S. et al. Nicastrin overexpression in transgenic mice induces aberrant behavior and APP processing. *Mol. Neurobiol.* **48**, 232–243 (2013).

- Francis, R. et al. *aph-1* and *pen-2* are required for Notch pathway signaling,  $\gamma$ -secretase cleavage of  $\beta$ APP, and presenilin protein accumulation. *Dev. Cell* **3**, 85–97 (2002).
- Takasugi, N. et al. The role of presenilin cofactors in the  $\gamma$ -secretase complex. *Nature* **422**, 438–441 (2003).
- Goutte, C., Tsunozaki, M., Hale, V. A. & Priess, J. R. APH-1 is a multipass membrane protein essential for the Notch signaling pathway in *Caenorhabditis elegans* embryos. *Proc. Natl Acad. Sci. USA* **99**, 775–779 (2002).
- Gu, Y. et al. APH-1 interacts with mature and immature forms of presenilins and nicastrin and may play a role in maturation of presenilin-nicastrin complexes. *J. Biol. Chem.* **278**, 7374–7380 (2003).
- Lu, P. et al. Three-dimensional structure of human  $\gamma$ -secretase. *Nature* **512**, 166–170 (2014).
- Li, X. et al. Structure of a presenilin family intramembrane aspartate protease. *Nature* **493**, 56–61 (2013).
- Xie, T. et al. Crystal structure of the  $\gamma$ -secretase component nicastrin. *Proc. Natl Acad. Sci. USA* **111**, 13349–13354 (2014).
- Sun, L. et al. Structural basis of human  $\gamma$ -secretase assembly. *Proc. Natl Acad. Sci. USA* **112**, 6003–6008 (2015).
- Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J. & Blundell, T. L. X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.* **214**, 199–222 (1990).
- Wang, J., Brunkan, A. L., Hecimovic, S., Walker, E. & Goate, A. Conserved “PAL” sequence in presenilins is essential for  $\gamma$ -secretase activity, but not required for formation or stabilization of  $\gamma$ -secretase complexes. *Neurobiol. Dis.* **15**, 654–666 (2004).
- Sato, C., Takagi, S., Tomita, T. & Iwatsubo, T. The C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the catalytic pore of the  $\gamma$ -secretase. *J. Neurosci.* **28**, 6264–6271 (2008).
- Wang, J. et al. C-terminal PAL motif of presenilin and presenilin homologues required for normal active site conformation. *J. Neurochem.* **96**, 218–227 (2006).
- Dang, S. et al. Cleavage of amyloid precursor protein by an archaeal presenilin homologue PSH. *Proc. Natl Acad. Sci. USA* **112**, 3344–3349 (2015).
- Schedin-Weiss, S., Winblad, B. & Tjernberg, L. O. The role of protein glycosylation in Alzheimer disease. *FEBS J.* **281**, 46–62 (2014).
- Pardossi-Piquard, R. et al. APH1 polar transmembrane residues regulate the assembly and activity of presenilin complexes. *J. Biol. Chem.* **284**, 16298–16307 (2009).
- Esselens, C. et al. Peptides based on the presenilin-APP binding domain inhibit APP processing and A $\beta$  production through interfering with the APP transmembrane domain. *FASEB J.* **26**, 3765–3778 (2012).
- Elad, N. et al. The dynamic conformational landscape of  $\gamma$ -secretase. *J. Cell Sci.* **128**, 589–598 (2015).
- Li, Y. et al. Structural interactions between inhibitor and substrate docking sites give insight into mechanisms of human PS1 complexes. *Structure* **22**, 125–135 (2014).
- Takagi-Niidome, S. et al. Cooperative roles of hydrophilic loop 1 and the C-terminus of presenilin 1 in the substrate-gating mechanism of  $\gamma$ -secretase. *J. Neurosci.* **35**, 2646–2656 (2015).
- Takeo, K., Watanabe, N., Tomita, T. & Iwatsubo, T. Contribution of the  $\gamma$ -secretase subunits to the formation of catalytic pore of presenilin 1 protein. *J. Biol. Chem.* **287**, 25834–25843 (2012).
- De Strooper, B. Lessons from a failed  $\gamma$ -secretase Alzheimer trial. *Cell* **159**, 721–726 (2014).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- DeLano, W. L. The PyMOL Molecular Graphics System (Schrödinger, LLC, 2002).

**Acknowledgements** We thank S. Chen and C. Savva for support with electron microscopy, and J. Grimmer and T. Darling for support with high-performance computing. This work was supported by funds from the Ministry of Science and Technology (2014ZX09507003006 to Y.S.), the National Natural Science Foundation of China (31130002 and 31321062 to Y.S.), a European Union Marie Curie Fellowship (to X.C.B.), and the UK Medical Research Council (MC\_UP\_A025\_1013, to S.H.W.S.).

**Author Contributions** Y.S. initiated and supervised the project. G.Y., P.L., D.M., L.S., and R.Z. prepared the sample and pre-screened samples in various detergents on F20. X.B. prepared grids and collected cryo-EM data. X.B. and S.S. calculated the cryo-EM map. C.Y. built and refined the atomic model. X.B. independently built and refined the atomic model. Y.S., L.S., G.Y., and R.Z. designed and analysed the mutational and biochemical characterizations. L.S., G.Y., and R.Z. performed the biochemical assays. All authors contributed to analysis of the structure. X.B., C.Y., S.S. and Y.S. contributed to manuscript preparation.

**Author Information** The atomic coordinates have been deposited in the Protein Data Bank under accession number 5A63, and the EM maps have been deposited in the Electron Microscopy Data Bank under accession code EMD-3061. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.B. (xcbai@mrc-lmb.cam.ac.uk), S.S. (scheres@mrc-lmb.cam.ac.uk), or Y.S. (shi-lab@tsinghua.edu.cn).



## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Protein preparation.** To improve protein production, we co-expressed the four components of  $\gamma$ -secretase, each in its own pMLink plasmid<sup>18</sup>. This strategy allowed both convenient manipulation of individual components and optimization of protein expression by altering the relative molar ratios of the four plasmids. Specific mutations in PS1, PEN-2, APH-1, or nicastrin were generated only in the corresponding pMLink vector using a standard PCR-based approach. PEN-2 has an amino-terminal Flag tag, APH-1 contains a carboxyl-terminal haemagglutinin (HA) tag, and nicastrin is tagged with a V5-His<sub>6</sub> sequence at the carboxyl terminus. The ten  $\gamma$ -secretase variants, each carrying a FAD-derived mutation on PS1, and the eight  $\gamma$ -secretase complexes, each carrying two introduced cysteine mutations, were overexpressed similarly as the WT  $\gamma$ -secretase<sup>18</sup>. Purification of both WT and mutant human  $\gamma$ -secretases followed published protocol<sup>18</sup>.

**Electron microscopy.** Aliquots of 3  $\mu$ l of purified  $\gamma$ -secretase in amphipols at a concentration of  $\sim 4 \mu$ M were placed on glow-discharged holey carbon grids (Quantifoil Au R1.2/1.3, 300 mesh), and flash frozen in liquid ethane using an FEI Vitrobot. Zero-energy-loss images were recorded manually on an FEI Titan Krios electron microscope at 300 kV, using a slit width of 20 eV on a GIF-Quantum energy filter. A Gatan K2-Summit detector was used in super-resolution counting mode at a calibrated magnification of  $\times 35,714$  (yielding a pixel size of 1.4 Å), and a dose rate of  $\sim 2.5$  electrons per square ångström per second (approximately 5 electrons per pixel per second). Exposures of 16 s were dose-fractionated into 20 movie frames. Defocus values in the final data set ranged from 0.7 to 3.2  $\mu$ m.

**Image processing.** We used MOTIONCORR<sup>37</sup> for whole-frame motion correction, CTFFIND3 (ref. 38) for estimation of the contrast transfer function parameters, and RELION-1.3 (ref. 39) for all other image processing steps. Templates for reference-based particle picking were obtained from 2D class averages that were calculated from a manually picked subset of the micrographs. Using low-pass filtered templates to 20 Å to limit reference bias, 1.8 million particles were picked automatically from a total of 2,925 micrographs. Because the picking procedure is prone to false positives<sup>40</sup>, we used reference-free 2D class averaging and an initial run of 3D classification to select 412,272 particles for a first 3D refinement. After per-particle motion correction and radiation-damage weighting (particle polishing<sup>41</sup>), these particles gave a reconstruction with a resolution of 3.5 Å. In a subsequent 3D classification, we applied a mask around the transmembrane domain and did not perform any alignments. This yielded a subset of 159,549 particles, for which a reconstruction with improved density in the trans-membrane domain was obtained. The resolution of this final map was 3.4 Å. Reported resolutions are based on the gold-standard FSC = 0.143 criterion<sup>39</sup>, and FSC curves were corrected for the effects of a soft mask on the FSC curve using high-resolution noise substitution<sup>42</sup>. All 3D classifications and refinements were started from a 40 Å low-pass filtered initial model, the first of which was made from our previous 4.5 Å map. Before visualization, all density maps were corrected for the modulation transfer function of the detector, and then sharpened by applying a negative B-factor that was estimated using automated procedures<sup>43</sup>. Local resolution variations were estimated using ResMap<sup>44</sup>.

**Atomic modelling.** The first atomic model was built from an intermediate, 3.9 Å EM map which was of sufficient quality to allow side-chain assignment for all four components of  $\gamma$ -secretase. The initial models used for nicastrin ECD and presenilin 1 were generated from the crystal structures of DpNCT<sup>20</sup> (Protein Data Bank accession number 4R12) and PSH<sup>19</sup> (Protein Data Bank accession number 4HYG) by CHAINSAW<sup>45</sup>. The two generated structures were first refined in real-space by PHENIX<sup>46</sup> with secondary structure and geometry restraints. APH-1 and PEN-2 were built *de novo* from a poly-Ala model. Then the entire atomic model was manually improved in COOT<sup>47</sup>. Sequence assignment was guided mainly by bulky residues such as Phe, Tyr, Trp and Arg. Unique patterns of sequences were exploited for validation of residue assignment. Eleven glycosylation sites with obvious sugar densities and five pairs of disulfide bonds in the nicastrin ECD domain also facilitated sequence assignment. This initial model

was largely confirmed and further improved by the 3.4 Å resolution EM map, with minor adjustment to a few residues and identification of two additional glycosylation sites and two lipid molecules. This model was refined in real-space by PHENIX<sup>46</sup>.

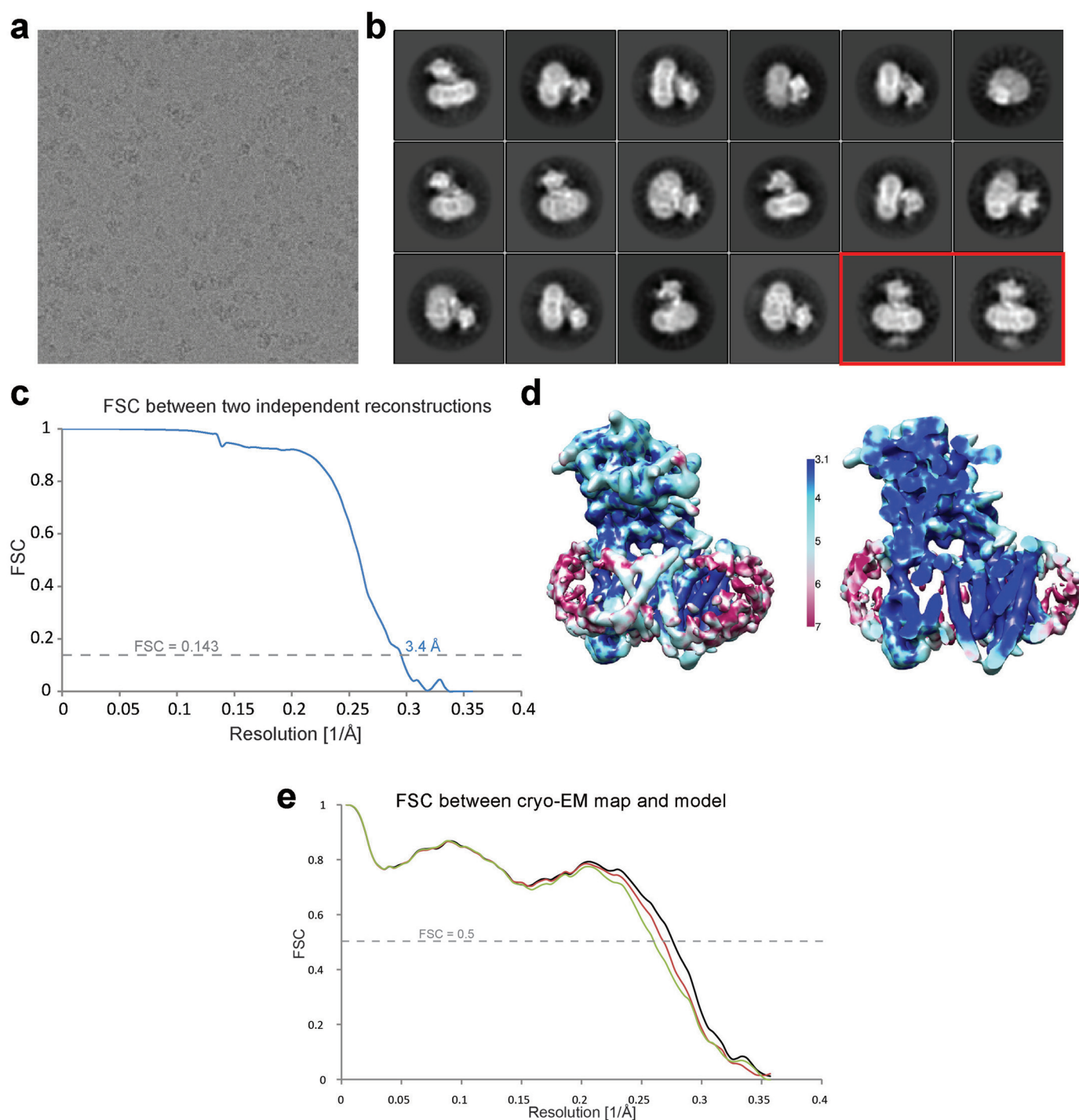
An independent effort of atomic model building used stereochemical refinement tools in Coot<sup>47</sup> and REFMAC<sup>48</sup> that were originally designed for X-ray crystallography, but later adapted for cryo-EM<sup>49</sup>. The structures of DpNCT<sup>20</sup> (Protein Data Bank accession number 4R12) and PSH<sup>19</sup> (Protein Data Bank accession number 4HYC) were used as a guide, while APH-1 and PEN-2 were built *de novo* in Coot<sup>47</sup>, starting from idealized  $\alpha$ -helices. This model was refined in REFMAC<sup>48</sup>, with secondary structure restraints generated by ProSMART<sup>50</sup>. Superposition of the two independently built models revealed that both models were in excellent agreement. Refinement of the final, consensus model was performed in REFMAC, and overfitting of the atomic model was monitored by refining the model in one of the two independent maps from the gold-standard refinement approach, and testing this model against the other map<sup>51</sup>. At no point in our procedures did we use the atomic model to modify the cryo-EM map.

**$\gamma$ -Secretase activity assay.** The assay was performed as described in the AlphaLISA Kit (PerkinElmer). Briefly, 2  $\mu$ l reaction products were incubated with 8  $\mu$ l AlphaLISA AB1-40/42 Acceptor beads at 23 °C for 1 h. After another incubation for 30 min with 10  $\mu$ l AlphaLISA AB1-40/42 Donor beads in the dark at 23 °C, the samples were read using Envision-Alpha Reader (PerkinElmer). The readings were expressed in arbitrary units.

**Crosslinking assay.** For the crosslinking assay, copper sulfate was added to the purified protein at a final concentration of 1 mM, and the sample was incubated at 4 °C for 1 h. Then, the protein was applied to gel filtration (Superose-6, GE Healthcare). The peak fractions were divided into two aliquots. One aliquot was treated with the reducing agent DL-dithiothreitol (DTT), at a final concentration of 10 mM and incubated at room temperature for 30 min. Samples from the two aliquots, treated with or without DTT, were analysed by western blotting using monoclonal antibodies against the HA tag on APH-1 or Flag tag on PEN-2. The antibodies were purchased from Beijing ComWin Biotech.

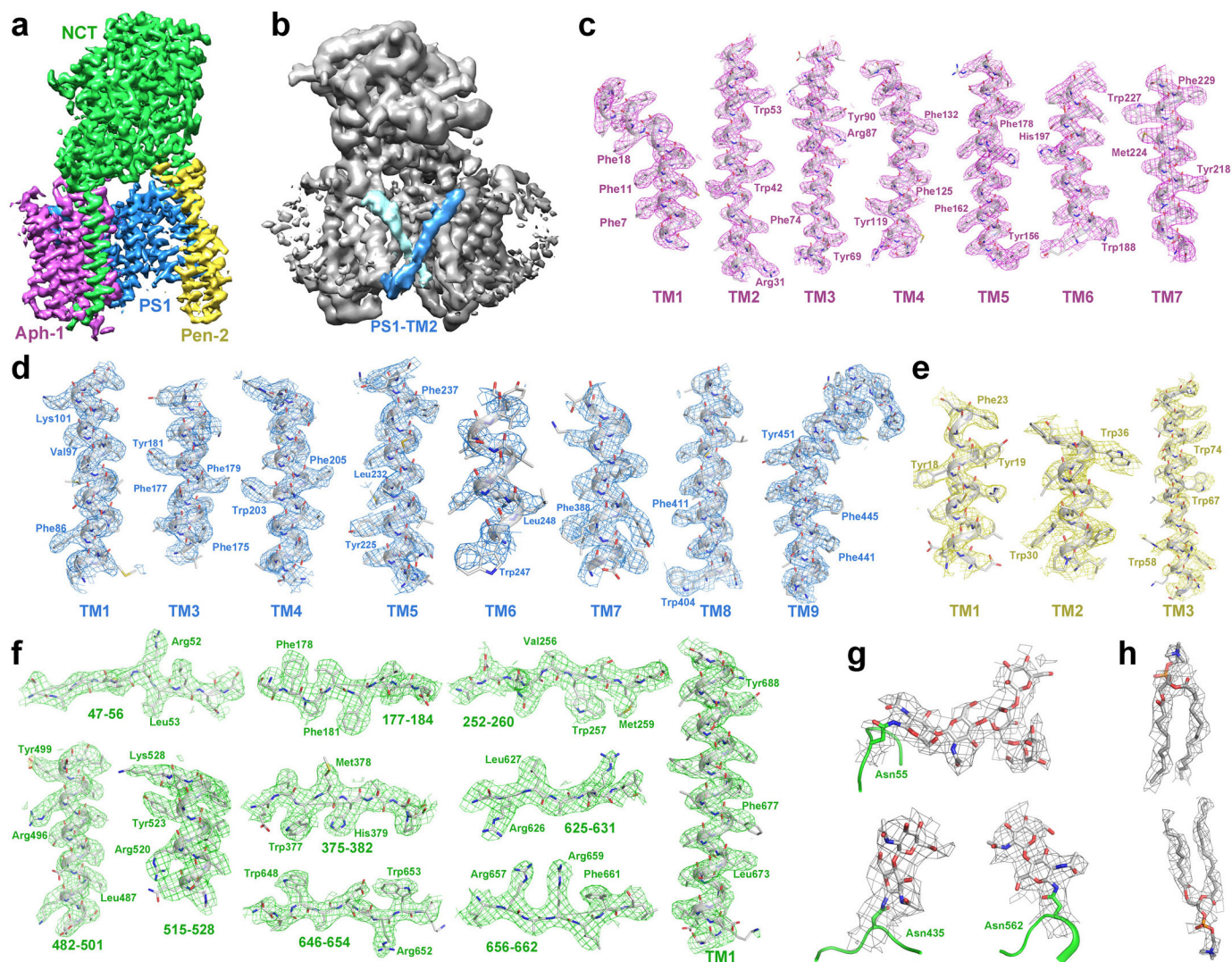
37. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
38. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
41. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
42. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
43. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
44. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
45. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* **41**, 641–643 (2008).
46. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
49. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
50. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).
51. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
52. Crystal, A. S. *et al.* Membrane topology of  $\gamma$ -secretase component PEN-2. *J. Biol. Chem.* **278**, 20117–20123 (2003).





**Extended Data Figure 1 | Cryo-EM, single-particle analysis of human  $\gamma$ -secretase.** **a**, Representative raw particles from an original micrograph. **b**, Representative reference-free 2D class averages of the  $\gamma$ -secretase particles. Two classes identified by a red rectangle box (lower right corner) may contain some density for the extended cytosolic loop sequences between TM6 and TM7 of PS1, which are disordered in the final maps. **c**, Resolution estimation of the EM structure. The overall resolution is calculated to be 3.4 Å on the basis of gold-standard FSC curve<sup>39</sup>. **d**, Colour-coded resolution variations in the

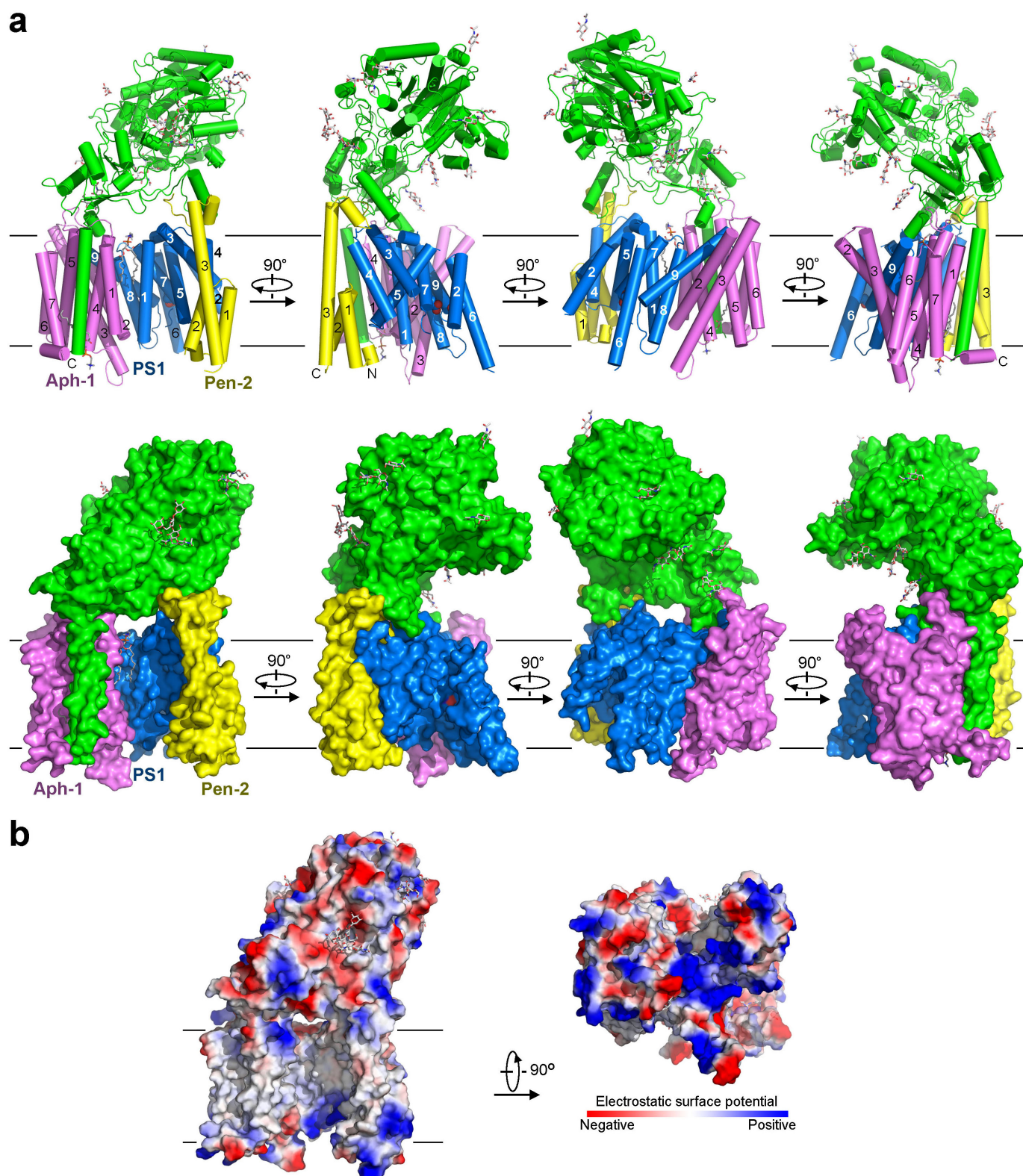
$\gamma$ -secretase structure as estimated by ResMap<sup>44</sup>. **e**, FSC curves of the final, Refmac-refined model versus the map it was refined against (in black); of a model refined in the first of the two independent maps used for the gold-standard FSC versus that same map (in red); and of a model refined in the first of the two independent maps versus the second independent map (in green). The small difference between the red and green curves indicates that the refinement of the atomic coordinates did not suffer from severe overfitting.



**Extended Data Figure 2 | An atomic model of human  $\gamma$ -secretase.** **a**, The  $\gamma$ -secretase structure is viewed parallel to the lipid membrane. Shown here is EM density for the entire  $\gamma$ -secretase complex. EM density is coloured blue for PS1, yellow for PEN-2, magenta for APH-1, and green for nicastrin. **b**, The density map for TM2 of PS1. Among the 20 TMs, TM2 of PS1 shows the highest degree of flexibility and only becomes visible at as rod-shaped density in a 7 Å low-pass filtered map. At this resolution, another rod-shaped density is visible next to TM2 and remains unaccounted for. **c**, EM density map and the

atomic model are shown for all seven TMs of APH-1. Two to three bulky residues are indicated for each TM. **d**, EM density map and the atomic model are shown for seven TMs of human PS1. TM6 exhibits relatively poor EM density, probably because of its intrinsic flexibility. **e**, EM density map and the atomic model are shown for the three TMs of PEN-2. **f**, EM density map and the atomic model are shown for the only TM and select regions of nicastrin. **g**, EM density map and the atomic model for three representative glycans. **h**, EM density map and putative assignment for two lipid molecules.



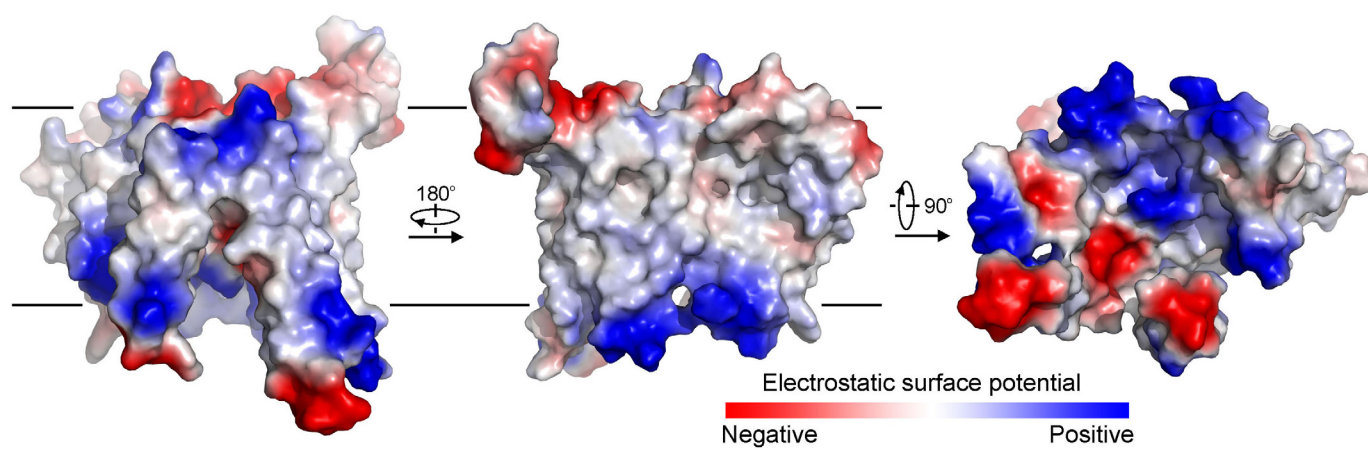


### Extended Data Figure 3 | Overall structure of human $\gamma$ -secretase.

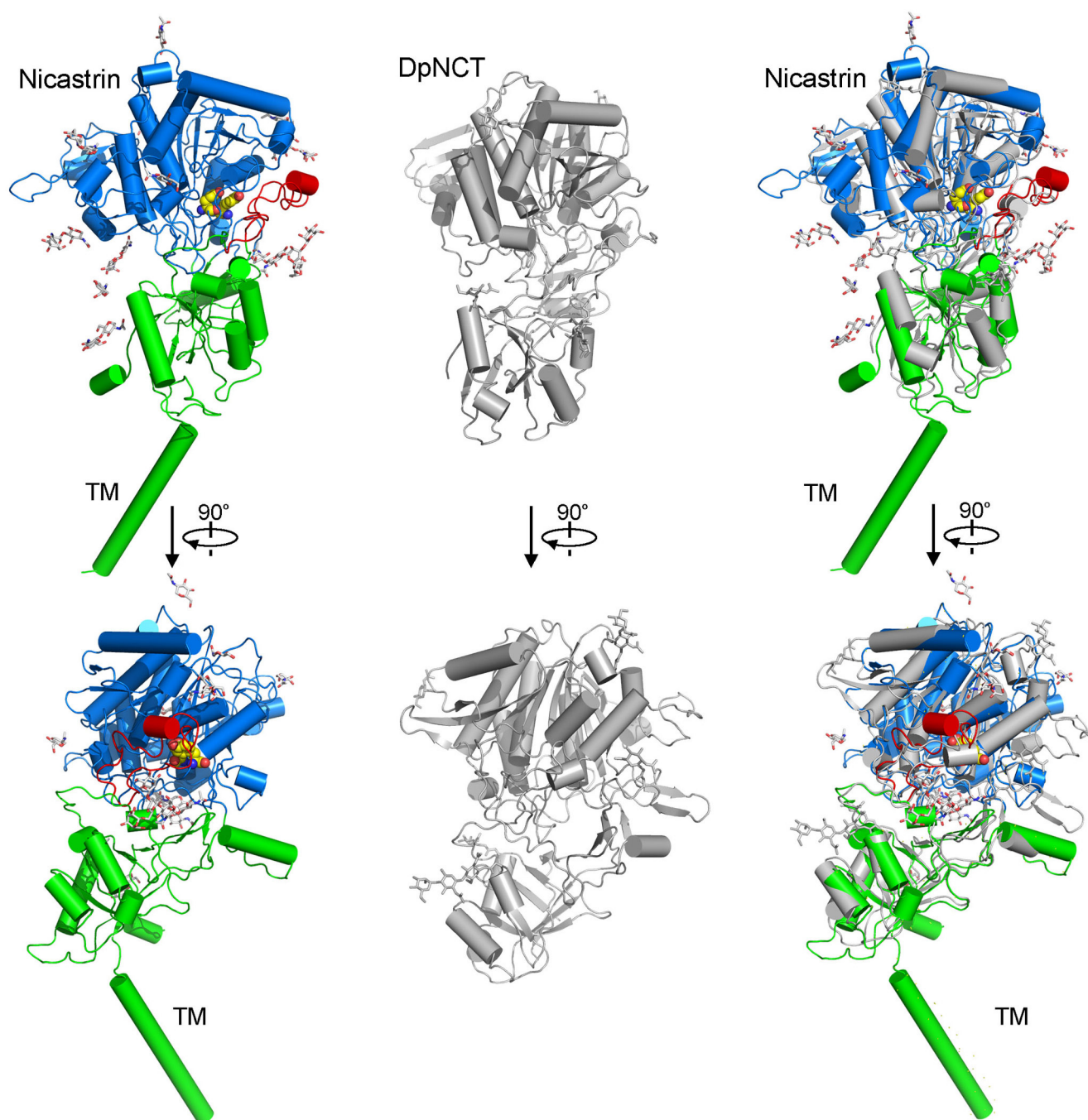
**a**, Structure of human  $\gamma$ -secretase is shown in cartoon representation (top) and surface view (bottom) in four successively perpendicular views. The  $\gamma$ -secretase structure is viewed parallel to the lipid membrane. The colouring scheme is

the same as in Fig. 1. Two lipid molecules are shown. Eleven glycosylated Asn residues and their glycans are displayed in stick. **b**, The  $\gamma$ -secretase structure is represented by electrostatic surface potential.



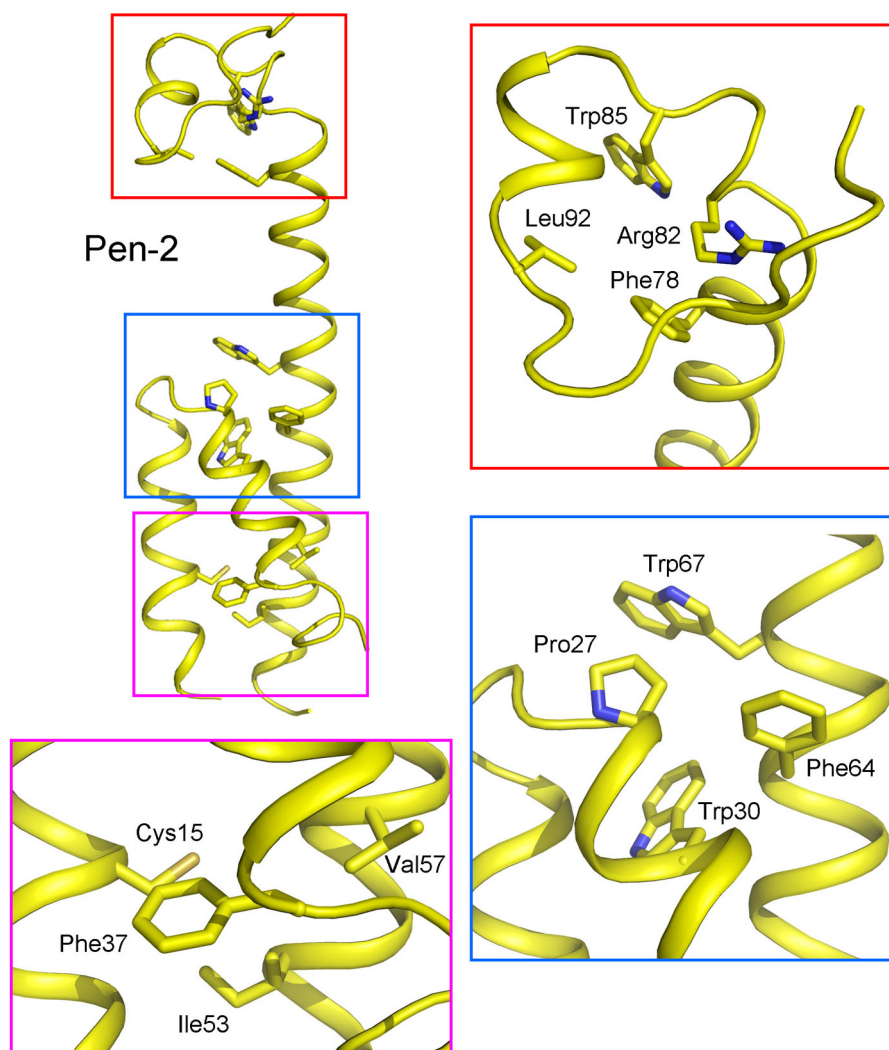


**Extended Data Figure 4 | Electrostatic surface potential of PS1.** PS1 exhibits a loosely folded structure, with several large cavities and empty spaces between adjacent TMs.



**Extended Data Figure 5 | Structural comparison between human nicastrin and nicastrin from *D. purpureum* (DpNCT).** Individual structures of human nicastrin and DpNCT<sup>20</sup> are shown in the left and middle panels,

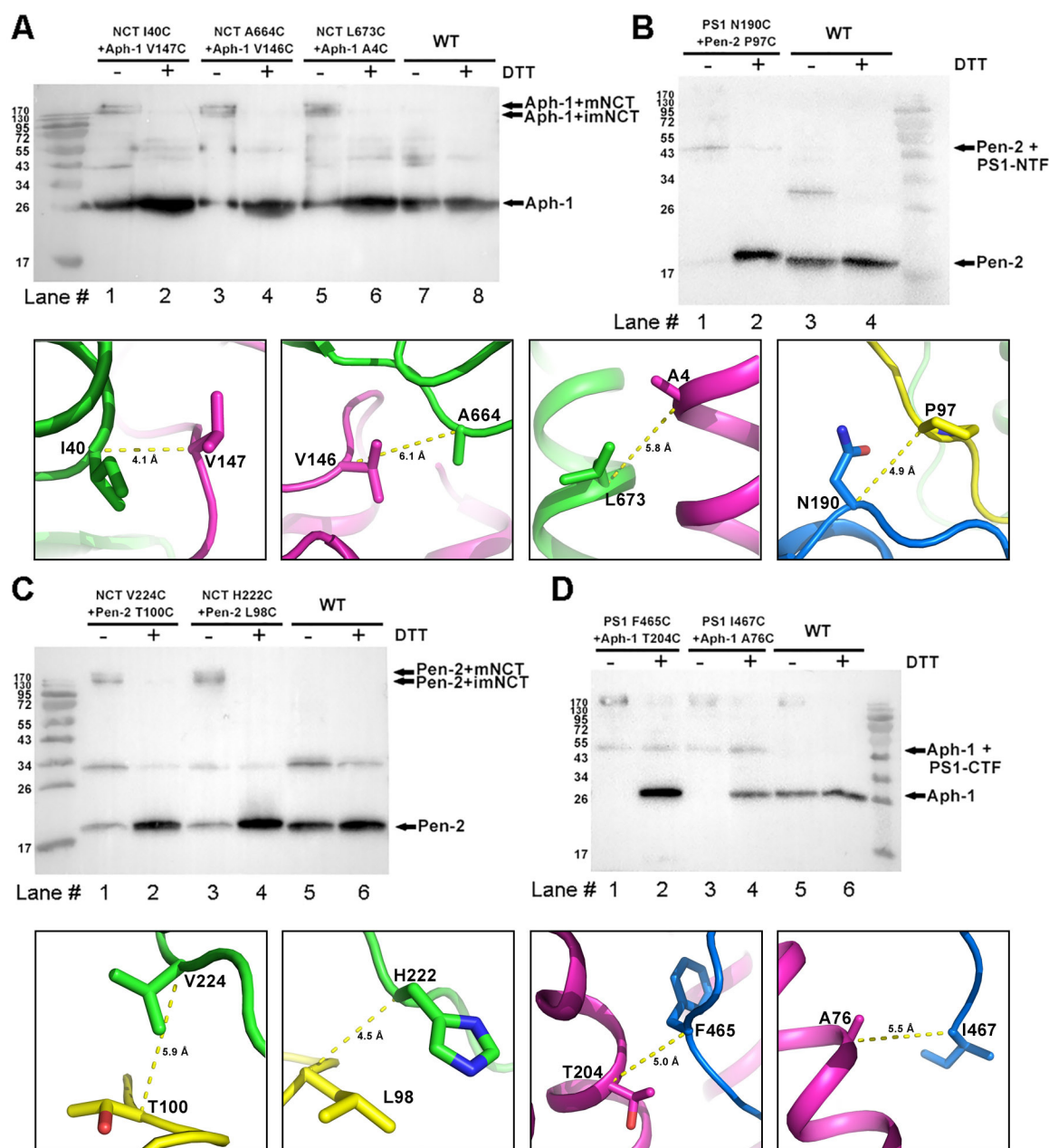
respectively. The overlay is shown in the right panel, with a root mean squared deviation of 2.2 Å. Two perpendicular views for each structure are displayed here.



**Extended Data Figure 6 | PEN-2 contains three small hydrophobic cores in its three TMs.** Unlike previous prediction<sup>14,52</sup>, PEN-2 contains three, not two, TMs. PEN-2 contains a small hydrophobic core in the extracellular side

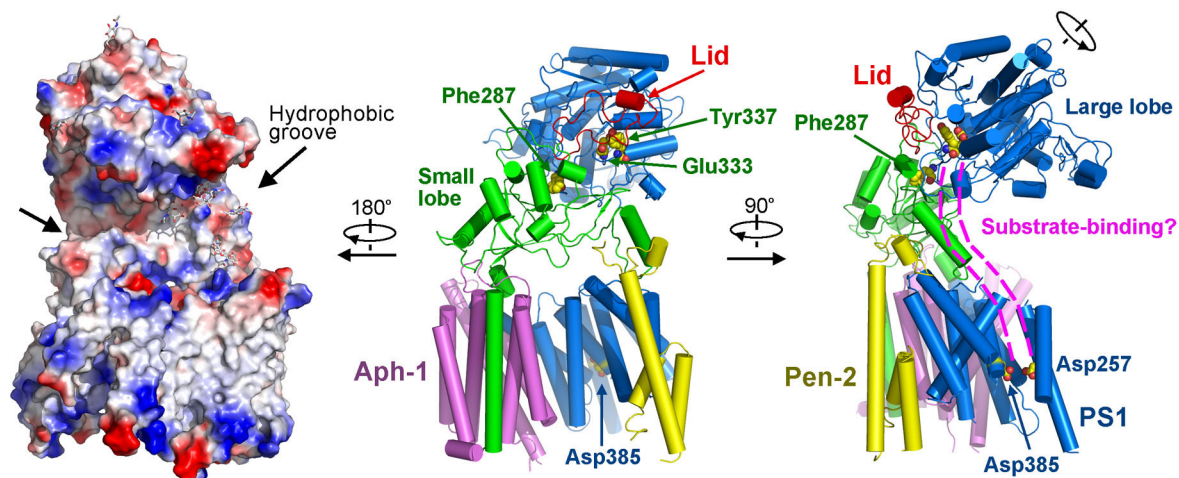
and two in the transmembrane region. These three regions are boxed and shown in close-up views.





**Extended Data Figure 7 | Results of crosslinking experiments corroborate the atomic model of  $\gamma$ -secretase.** **a**, Crosslinking results for the interface between APh-1 and nicastrin (NCT). Three mutant  $\gamma$ -secretase complexes were examined: APh-1-V147C/NCT-I40C, APh-1-V146C/NCT-A664C, and APh-1-A4C/NCT-L673C. Shown in the upper panel is an SDS-PAGE gel blotted by a monoclonal antibody against the HA tag on APh-1. Only in the absence of DTT did crosslinking lead to high molecular mass complexes for the mutant  $\gamma$ -secretase, but not for the WT  $\gamma$ -secretase. The two bands probably represent APh-1 crosslinked to mature nicastrin (mNCT) and immature nicastrin (iNCT). The structural basis is shown in the lower panel. The

distances between the C $\alpha$  atoms of the two residues targeted for cysteine mutation range from 4.1 to 6.1 Å, which would facilitate convenient crosslinking reactions. **b**, Crosslinking results for the interface between PS1 and PEN-2. The mutant  $\gamma$ -secretase contains PEN-2-P97C and PS1-N190C. Shown in the upper panel is an SDS-PAGE gel blotted by a monoclonal antibody against the Flag tag on PEN-2. **c**, Crosslinking results for the interface between PEN-2 and nicastrin. Two  $\gamma$ -secretase mutants were examined: PEN-2-T100C/NCT-V224C and PEN-2-L98C/NCT-H222C. **d**, Crosslinking results for the interface between APh-1 and PS1. Two  $\gamma$ -secretase mutants were examined: APh-1-T204C/PS1-F465C and APh-1-A76C/PS1-I467C.



**Extended Data Figure 8 | Implication on substrate access to  $\gamma$ -secretase.** Structure of  $\gamma$ -secretase is displayed in three relevant views: left, electrostatic surface potential from the convex side of  $\gamma$ -secretase; middle, overall structure,

with key features labelled; right, suggested putative path for substrate access to the active site of  $\gamma$ -secretase.

Extended Data Table 1 | Refinement and model statistics

<b>Data Collection</b>	
Particles	159,549
Pixel size (Å)	1.4
Defocus range (μm)	0.7-3.2
Voltage (kV)	300
Electron dose (e-/Å <sup>-2</sup> )	40
<b>Model composition</b>	
Non-hydrogen atoms	10,016
Protein residues	1,251
Phospholipids (PC)	2
Sugars	15
<b>Refinement</b>	
Resolution (Å)	3.4
Map sharpening B-factor (Å <sup>2</sup> )	-109
Fourier Shell Correlation	0.8503
Rfactor	0.3494
<b>Rms deviations</b>	
Bonds (Å)	0.0077
Angles (°)	1.4224
<b>Validation (proteins)</b>	
Molprobity score	2.65
Clashscore, all atoms	8.22
Good rotamers (%)	93.5%
<b>Ramachandran plot</b>	
Favored (%)	88.5
Allowed (%)	8.6
Outliers (%)	2.9



# Dense cloud cores revealed by CO in the low metallicity dwarf galaxy WLM

Monica Rubio<sup>1</sup>, Bruce G. Elmegreen<sup>2</sup>, Deidre A. Hunter<sup>3</sup>, Elias Brinks<sup>4</sup>, Juan R. Cortés<sup>5,6</sup> & Phil Cigan<sup>7</sup>

Understanding stellar birth requires observations of the clouds in which they form. These clouds are dense and self-gravitating, and in all existing observations they are molecular, with H<sub>2</sub> the dominant species and carbon monoxide (CO) the best available tracer<sup>1,2</sup>. When the abundances of carbon and oxygen are low compared with that of hydrogen, and the opacity from dust is also low, as in primeval galaxies and local dwarf irregular galaxies<sup>3</sup>, CO forms slowly and is easily destroyed, so it is difficult for it to accumulate inside dense clouds<sup>4</sup>. Here we report interferometric observations of CO clouds in the local group dwarf irregular galaxy Wolf–Lundmark–Melotte (WLM)<sup>5</sup>, which has a metallicity that is 13 per cent of the solar value<sup>6,7</sup> and 50 per cent lower than the previous CO detection threshold. The clouds are tiny compared to the surrounding atomic and H<sub>2</sub> envelopes, but they have typical densities and column densities for CO clouds in the Milky Way. The normal CO density explains why star clusters forming in dwarf irregulars have similar densities to star clusters in giant spiral galaxies. The low cloud masses suggest that these clusters will also be low mass, unless some galaxy-scale compression occurs, such as an impact from a cosmic cloud or other galaxy. If the massive metal-poor globular clusters in the halo of the Milky Way formed in dwarf galaxies, as is commonly believed, then they were probably triggered by such an impact.

WLM is an isolated dwarf galaxy at a distance of  $985 \pm 33$  kiloparsecs (kpc) (ref. 5). Like other dwarfs, the relative abundance of supernova-processed elements (‘metals’) such as carbon and oxygen is low<sup>6</sup>,  $12 + \log(\text{O}/\text{H}) = 7.8$ , compared with 8.66 for the Milky Way<sup>7</sup>. Low carbon and oxygen abundances, along with the correspondingly low abundances of other processed elements and dust, make the CO molecule rare compared to H<sub>2</sub>, and this calls into question the standard model of star formation in CO-rich clouds<sup>1</sup>. In fact, the star formation rate<sup>8</sup> compared with the existing stellar mass is actually high in WLM:  $0.006 M_{\odot} \text{ yr}^{-1}$  of new stars for a total stellar mass<sup>9</sup> of  $1.6 \times 10^7 M_{\odot}$  is 12 times higher than in the Milky Way, where the star formation rate<sup>10</sup> is  $\sim (1.9 \pm 0.4) M_{\odot} \text{ yr}^{-1}$  and the stellar mass is  $(6.4 \pm 0.6) \times 10^{10} M_{\odot}$  (ref. 11). Thus, WLM forms stars efficiently even with a relatively low abundance of CO.

To understand star formation in metal-poor galaxies, which include the most numerous galaxies in the local universe (the dwarfs, plus all primeval galaxies), we previously searched for CO(3–2) in WLM using the APEX telescope<sup>12</sup>, discovering it in two unresolved regions at an abundance relative to H<sub>2</sub> that was half that in the next-lowest metallicity galaxy, the Small Magellanic Cloud. Now, with the completion of the new millimetre- and submillimetre-wavelength interferometer, the Atacama Large Millimeter Array (ALMA), we have imaged these two regions in CO(1–0) and resolved the actual molecular structure.

The ALMA maps with  $6.2 \times 4.3$  pc spatial resolution (HPBW), 5 mJy sensitivity and  $0.5 \text{ km s}^{-1}$  velocity resolution (FWHM) contain 10 CO clouds with an average radius of 2 pc and an average virial mass

of  $2 \times 10^3 M_{\odot}$ . Figure 1 shows the CO emission with black contours superposed on H I in green and H $\alpha$  in red. The inset shows a colour composite of the optical image in green (V band), the far ultraviolet (FUV) Galaxy Evolution Explorer (GALEX) image in blue and the H I in red. A [C II] $\lambda 158 \mu\text{m}$  image from the Herschel Space Observatory<sup>13</sup> is superposed on the southeast region in blue<sup>14</sup>. The [C II] is from a photodissociation region including ionized carbon; it is five times larger in size than the CO core, indicating a gradual transition between low-density atomic gas to high-density molecular gas.

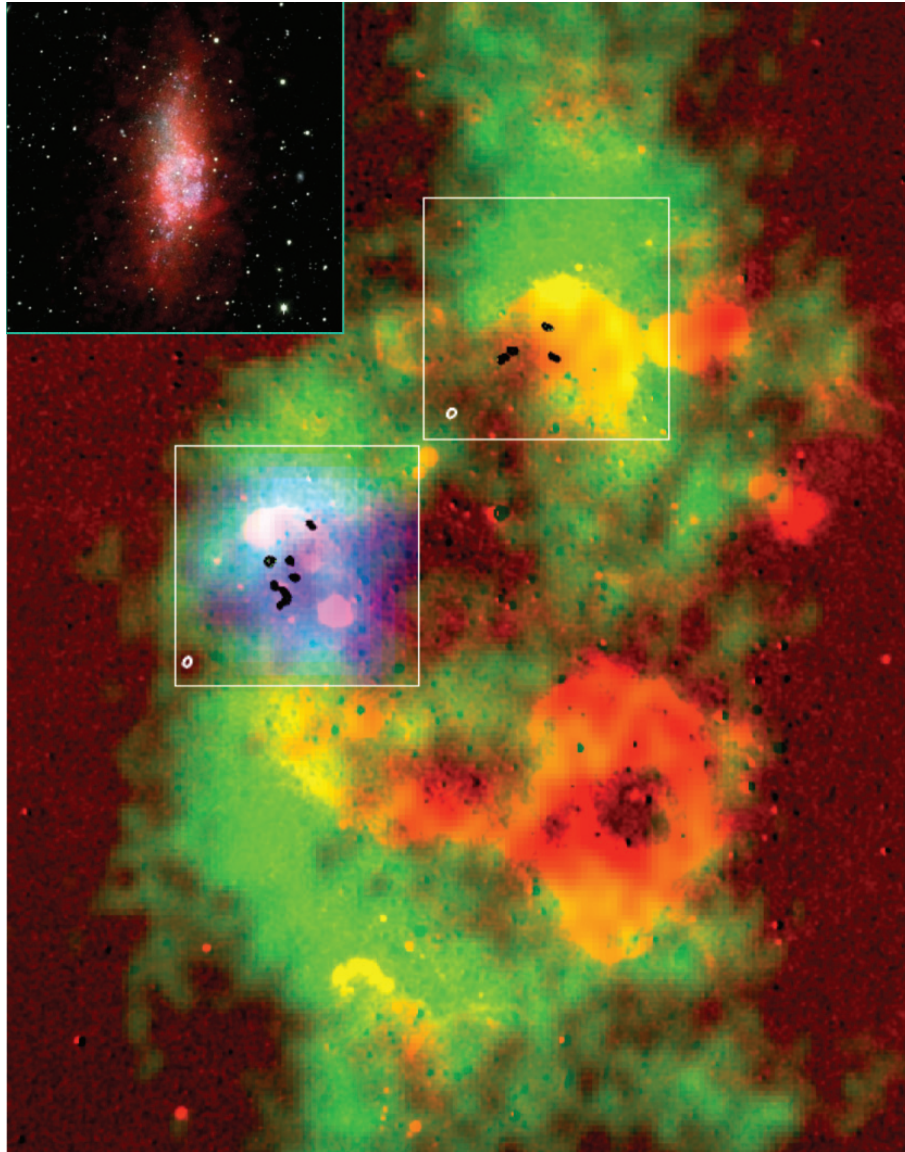
Figure 2 shows the contours and spectra of each cloud. The spectral signal-to-noise averages 10 when smoothed to the typical linewidth of  $2 \text{ km s}^{-1}$ . Velocities for H I emission are indicated by a bar below each CO spectrum. The cloud properties are summarized in Table 1. The radii  $R$  range from 1.5 pc to 6 pc, obtained using the equation  $R = (A/\pi)^{0.5}$  for area  $A$ , with  $A$  determined after deconvolution by quadratic difference with the beam area. The sum of all the line emissions measured by ALMA is within a factor of two of the total emission found at  $18''$  resolution by the APEX telescope. The linewidths were corrected for instrumental spectral broadening.

Virial masses for the CO clouds were calculated from the relation  $M_{\text{vir}}(M_{\odot}) = 1,044 R \sigma^2$  for  $R$  in pc and Gaussian linewidths  $\sigma$  in  $\text{km s}^{-1}$ . The CO luminosity in  $\text{K km s}^{-1} \text{ pc}^2$  was calculated from  $L_{\text{CO}} = 2,453 S_{\text{CO}} \Delta V D^2$  for integrated emission  $S$  in  $\text{Jy km s}^{-1}$ , FWHM of the line  $\Delta V$  in  $\text{km s}^{-1}$ , and distance  $D$  in Mpc. Figure 3 shows the relationships between these values including other dwarf galaxies (all for CO(1–0)). The CO clouds in WLM satisfy the usual correlations, although they are the smallest seen for any of these galaxies. Higher-resolution observations should reveal small clouds and/or cores in other galaxies too, but the main point is that WLM has no CO clouds as large as those seen elsewhere.

The virial mass gives some perspective on the conversion from CO luminosity to mass derived previously<sup>12</sup>, which was  $\alpha_{\text{CO}} \approx (124 \pm 60) M_{\odot} \text{ pc}^{-2} (\text{K km s}^{-1})^{-1}$  for the northwest region. This value for  $\alpha$  was derived from the dust-derived H<sub>2</sub> column density. If instead we take the virial masses and CO luminosities in Table 1, we find that the mean ratio is  $\alpha_{\text{vir}} \approx (28 \pm 28) M_{\odot} \text{ pc}^{-2} (\text{K km s}^{-1})^{-1}$ . If the clouds are not gravitationally bound, then  $\alpha_{\text{vir}}$  would be smaller. The difference between these two  $\alpha$  values arises because most of the H<sub>2</sub> volume has no CO emission, which apparently exists only in the densest cores of the H<sub>2</sub> clouds. For the Milky Way, CO and H<sub>2</sub> have about the same extent in star-forming clouds, making  $\alpha_{\text{CO}} \approx 4 M_{\odot} \text{ pc}^{-2} (\text{K km s}^{-1})^{-1}$ . When CO does not fill an H<sub>2</sub> cloud,  $\alpha$  can be small for each CO core but large for the total H<sub>2</sub> cloud. If the purpose of  $\alpha$  is to determine the total H<sub>2</sub> mass in a region based on  $L_{\text{CO}}$ , then the large value should be used.

The self-gravitational boundedness of the CO clouds can be estimated from the general requirement of an associated H<sub>2</sub> density of  $\sim 10^3 \text{ cm}^{-3}$  for collisional excitation<sup>15</sup>. In fact, the virial density of the CO clouds is comparable to this,  $n(\text{H}_2) = 4.1 \times 10^{-21} \text{ g cm}^{-3}$

<sup>1</sup>Departamento de Astronomía, Universidad de Chile, Casilla 36-D, 8320000 Santiago, Chile. <sup>2</sup>IBM Research Division, T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, New York 10598, USA. <sup>3</sup>Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, Arizona 86001, USA. <sup>4</sup>Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK. <sup>5</sup>Joint ALMA Observatory, Alonso de Córdova 3107, Vitacura, 7630355 Santiago, Chile. <sup>6</sup>National Radio Astronomy Observatory, Avenida Nueva Costanera 4091, Vitacura, 7630197 Santiago, Chile. <sup>7</sup>New Mexico Institute of Mining and Technology, Socorro, New Mexico 87801, USA.



**Figure 1 | Tiny CO clouds in WLM.** A colour composite of the various gas phases in WLM: green is H I (ref. 26), red is H $\alpha$  (ref. 27) and blue is [C II] $\lambda$ 158  $\mu$ m (ref. 14). The CO emission is shown as black single contours inside the 1 arcmin  $\times$  1 arcmin white squares that outline the area mapped in

$^{12}\text{CO}$  (1-0) by ALMA. The synthesized ALMA beam ( $0.9'' \times 1.3''$ ) is shown in the bottom left corner of each square. The inset in the top left is the full view of WLM obtained by combining H I and optical data: red is H I, green is V band, and blue is GALEX FUV<sup>26</sup>.

( $\sim 10^3 \text{ cm}^{-3}$ ), from the ratio of the virial mass ( $\sim 2 \times 10^3 M_{\odot}$ ) to the cloud volume ( $4\pi R^3/3$  for  $R \approx 2 \text{ pc}$ ). Thus the clouds could be marginally bound.

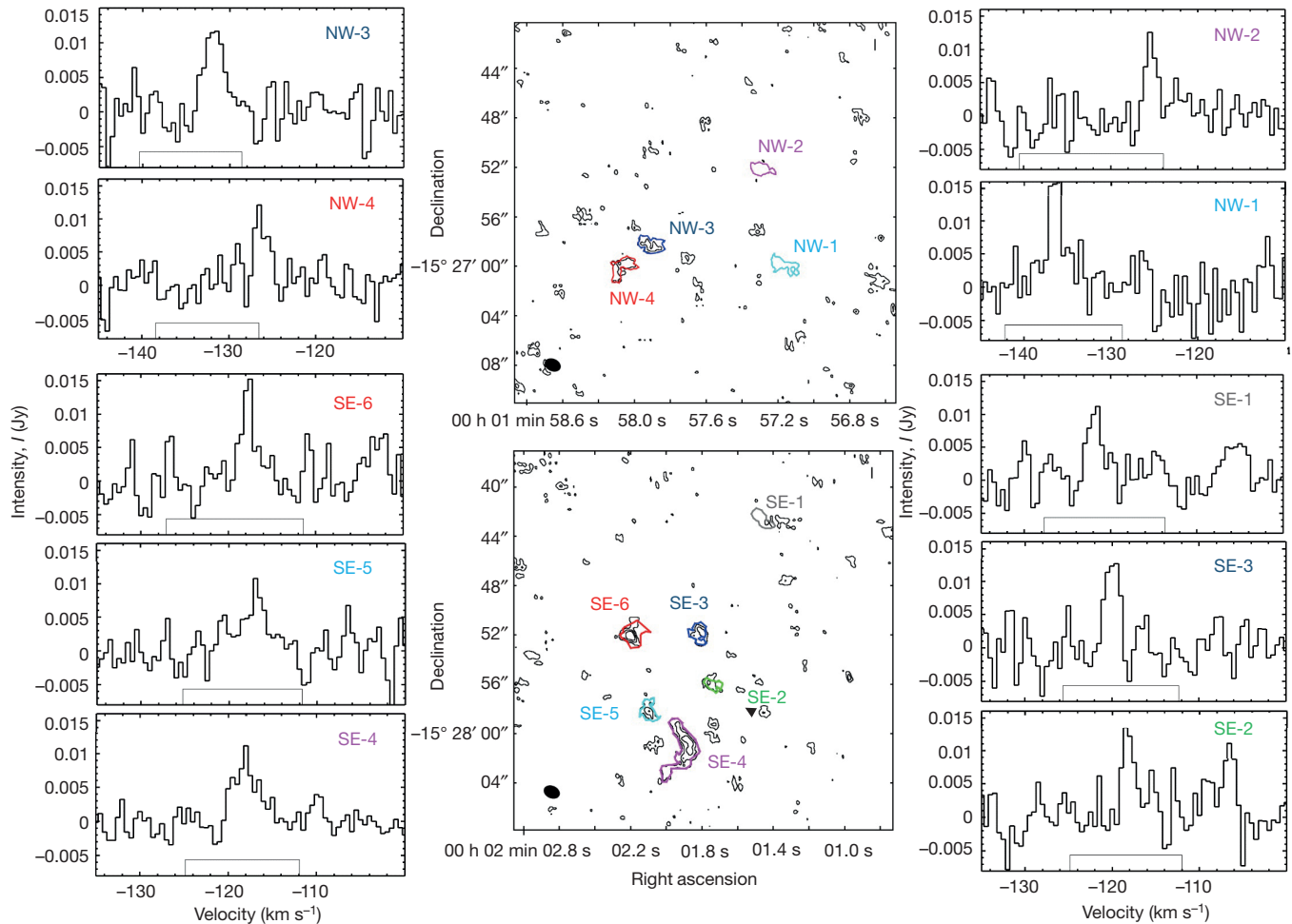
Another measure of CO density is from pressure equilibrium between the CO regions and the weight of the overlying H I and H $_2$  layers. The H $_2$  mass column density,  $\Sigma_{\text{H}_2}$ , comes from the difference

between the total gas column density derived from the dust emission and the H I column density observed at 21 cm. For the northwest region<sup>12</sup>,  $\Sigma_{\text{H}_2} = (31 \pm 15) M_{\odot} \text{ pc}^{-2}$ . Adding the H I column density<sup>12</sup> gives  $\Sigma_{\text{total}} = (58 \pm 15) M_{\odot} \text{ pc}^{-2}$ . The corresponding pressure from self-gravity is  $(\pi/2)G\Sigma_{\text{total}}^2 \approx 1.6 \times 10^{-11} \text{ dynes}$ . Considering the typical CO velocity dispersion for our clouds of  $\sigma \approx 0.9 \text{ km s}^{-1}$ , the ratio

**Table 1 | Properties of WLM CO clouds**

Region	RA (h min s)	Dec. ( $^{\circ}$ ' ")	Peak intensity (mJy)	$V_{\text{LSR}}$ (km s $^{-1}$ )	Integrated flux density (Jy km s $^{-1}$ )	Radius (pc)	$\sigma$ (km s $^{-1}$ )	$M_{\text{vir}} (M_{\odot})$	$L_{\text{CO}}$ (K km s $^{-1}$ pc $^2$ )
NW-1	00 01 57.162	-15 27 00.00	12.2	$-131.79 \pm 0.19$	$0.037 \pm 0.004$	$2.21 \pm 1.11$	$1.05 \pm 0.17$	$2,548 \pm 1,522$	$81.47 \pm 8.39$
NW-2	00 01 57.291	-15 26 52.80	16.1	$-136.42 \pm 0.18$	$0.025 \pm 0.003$	$1.49 \pm 0.77$	$0.84 \pm 0.28$	$1,087 \pm 919$	$56.23 \pm 5.69$
NW-3	00 01 57.901	-15 26 58.00	10.8	$-126.27 \pm 0.15$	$0.048 \pm 0.005$	$2.69 \pm 1.35$	$0.75 \pm 0.14$	$1,561 \pm 985$	$106.26 \pm 10.71$
NW-4	00 01 58.079	-15 27 00.12	12.2	$-125.38 \pm 0.16$	$0.025 \pm 0.003$	$2.69 \pm 1.35$	$0.57 \pm 0.14$	$898.4 \pm 637$	$54.90 \pm 5.84$
SE-1	00 02 01.485	-15 27 42.65	10.8	$-121.85 \pm 0.18$	$0.051 \pm 0.005$	$1.68 \pm 0.87$	$0.77 \pm 0.18$	$1,037 \pm 720$	$113.57 \pm 11.46$
SE-2	00 02 01.761	-15 27 55.83	13.3	$-118.18 \pm 0.16$	$0.021 \pm 0.002$	$<1$	$0.61 \pm 0.23$	$<390 \pm 300$	$46.93 \pm 5.02$
SE-3	00 02 01.801	-15 27 51.78	14.3	$-120.00 \pm 0.12$	$0.030 \pm 0.003$	$2.21 \pm 1.15$	$0.69 \pm 0.09$	$1,113 \pm 653$	$67.30 \pm 6.96$
SE-4	00 02 01.864	-15 28 00.52	8.77	$-118.01 \pm 0.17$	$0.258 \pm 0.026$	$6.01 \pm 1.20$	$1.32 \pm 0.14$	$10,881 \pm 3,209$	$571.17 \pm 57.20$
SE-5	00 02 02.101	-15 27 58.23	6.92	$-117.21 \pm 0.48$	$0.030 \pm 0.003$	$2.02 \pm 0.96$	$1.81 \pm 0.57$	$6,896 \pm 5,426$	$67.30 \pm 7.16$
SE-6	00 02 02.222	-15 27 52.08	13.7	$-117.79 \pm 0.12$	$0.031 \pm 0.003$	$3.37 \pm 1.06$	$0.63 \pm 0.15$	$1,383 \pm 805$	$68.85 \pm 7.11$

Dec., declination; NW, northwest; RA, right ascension; SE, southeast.



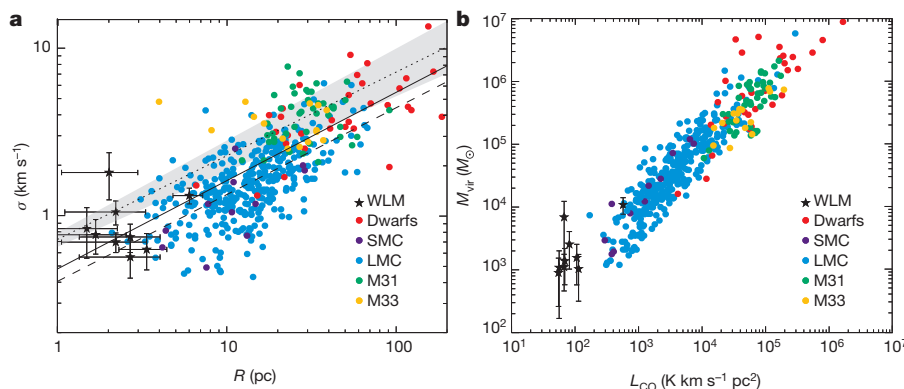
**Figure 2 | CO clouds and spectra.** CO contour maps of the integrated emission starting at the 2-s.d. level (right ascension and declination in J2000.0 coordinates). Different CO clouds are identified by colour. The ALMA beam is the black ellipse in the lower left corner. The CO spectrum corresponding to

each detection is plotted. The velocity for H I emission (FWHM) is shown as a rectangular box on the abscissa (local standard of rest, LSR); the CO velocities agree with the H I.

of the core pressure to the square of the CO velocity dispersion is the equilibrium core density,  $1.9 \times 10^{-21} \text{ g cm}^{-3}$ , corresponding to  $500 \text{ H}_2 \text{ cm}^{-3}$ . Thus the virial density, excitation density and pressure equilibrium density are all about  $10^3 \text{ cm}^{-3}$ .

A condition for molecules in the Milky Way is a threshold extinction of  $A_V = 0.3 \text{ mag}$  for  $\text{H}_2$  and  $\sim 1.5 \text{ mag}$  for  $\text{CO}^{16}$ . These correspond to mass column densities of  $6.1 M_\odot \text{ pc}^{-2}$  and  $30.3 M_\odot \text{ pc}^{-2}$  in the solar neighbourhood. In WLM, where the metallicity is 13% solar, the mass thresholds are  $47 M_\odot \text{ pc}^{-2}$  and  $230 M_\odot \text{ pc}^{-2}$  for the same extinctions, respectively. The first is satisfied by the H I +  $\text{H}_2$  envelope

of the CO cores ( $\sim 58 M_\odot \text{ pc}^{-2}$ ), and the second is satisfied by the total column density of  $220 M_\odot \text{ pc}^{-2}$  calculated from the H I and  $\text{H}_2$  envelope, plus the  $\text{H}_2$  from the embedded CO core itself (as determined from the CO virial mass,  $2 \times 10^3 M_\odot$ , and ALMA measured radius, 2 pc). These results suggest that the CO clouds in WLM are normal in terms of density, pressure and column density, which explains why they lie on the standard correlations in Fig. 3. They also appear to be marginally self-bound by gravity, suggesting that they are related to star formation. Their properties are typical for parsec-size molecular cloud cores in the solar neighbourhood<sup>17</sup>.



**Figure 3 | Correlations for CO clouds in dwarf galaxies.** **a, b.** The symbols refer to different galaxies (SMC, dwarfs, M31 and M33 (ref. 28); LMC (ref. 29)). Error bars are 1 s.d. **a.** CO linewidth  $\sigma$  versus radius  $R$ ; the solid line is a fit to WLM, the SMC and dwarf galaxies:  $\sigma = (0.48 \pm 0.08) \times R^{0.53 \pm 0.05}$ ; and the dashed line includes the LMC also:  $\sigma = (0.40 \pm 0.03) \times R^{0.52 \pm 0.05}$ . The black short-dashed line and the grey area indicate the standard relation for the Milky Way<sup>30</sup>:  $\sigma = (0.72 \pm 0.07) \times R^{0.50 \pm 0.05}$ .  $R$  for WLM is measured in the same way as for the Milky Way and other galaxies. **b.** Virial mass versus CO luminosity.



Our observation explains why star clusters have about the same central densities in dwarf irregular<sup>18</sup> and spiral galaxies<sup>19</sup>, even though the ambient gas density in dwarfs is much less than it is in spirals<sup>20</sup>. If the unifying process of star formation is the formation of CO and other asymmetric molecules for cooling (however, see refs 16, 21), then the similarity between the CO cores in the two cases accounts for the uniformity of clusters. The small mass of the CO cores in WLM also explains why most dwarf galaxies do not form high-mass clusters<sup>18</sup>. The CO parts of interstellar clouds are smaller at lower metallicities, so the clusters that result are smaller too. For example, there are no massive young clusters in these regions of WLM<sup>18</sup>. This lack of massive clusters is usually attributed to sparse sampling of the cluster mass distribution function at low star formation rates<sup>18</sup>, but the present observations suggest it could result from some physical reason too, such as the lack of massive CO clouds at low metallicity.

When the local dwarf galaxies NGC 1569 and NGC 5253 formed massive clusters, there was a major impact event to increase the pressure and mass at high density<sup>22,23</sup>. Such an impact would also seem to be needed for the formation of old halo globular clusters, which are massive and have low metallicity like their former dwarf galaxy hosts<sup>24,25</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 April; accepted 10 July 2015.

- McKee, C. F. & Ostriker, E. C. Theory of star formation. *Annu. Rev. Astron. Astrophys.* **45**, 565–687 (2007).
- Kennicutt, R. C. & Evans, N. J. Star formation in the Milky Way and nearby galaxies. *Annu. Rev. Astron. Astrophys.* **50**, 531–608 (2012).
- R  my-Ruyer, A. *et al.* Revealing the cold dust in low-metallicity environments. I. Photometry analysis of the Dwarf Galaxy Survey with Herschel. *Astron. Astrophys.* **557**, A95 (2013).
- Beuther, H. Carbon in different phases ([CII], [CI], and CO) in infrared dark clouds: cloud formation signatures and carbon gas fractions. *Astron. Astrophys.* **571**, A53 (2014).
- Leaman, R. *et al.* The resolved structure and dynamics of an isolated dwarf galaxy: a VLT and Keck spectroscopic survey of WLM. *Astrophys. J.* **750**, 33 (2012).
- Lee, H., Skillman, E. D. & Venn, K. A. Investigating the possible anomaly between nebular and stellar oxygen abundances in the dwarf irregular galaxy WLM. *Astrophys. J.* **620**, 223–237 (2005).
- Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
- Hunter, D. A., Elmegreen, B. G. & Ludka, B. C. GALEX ultraviolet imaging of dwarf galaxies and star formation rates. *Astron. J.* **139**, 447–475 (2010).
- Zhang, H.-X., Hunter, D. A., Elmegreen, B. G., Gao, Y. & Schruha, A. Outside-in shrinking of the star-forming disks of dwarf irregular galaxies. *Astron. J.* **143**, 47 (2012).
- Chomiuk, L. & Povich, M. S. Toward a unification of star formation rate determinations in the Milky Way and other galaxies. *Astron. J.* **142**, 197 (2011).
- McMillan, P. J. Mass models of the Milky Way. *Mon. Not. R. Astron. Soc.* **414**, 2446–2457 (2011).
- Elmegreen, B. G. *et al.* Carbon monoxide in clouds at low metallicity in the dwarf irregular galaxy WLM. *Nature* **495**, 487–489 (2013).
- Pilbratt, G. L. *et al.* Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy. *Astron. Astrophys.* **518**, L1 (2010).
- Cigan, P. *et al.* Herschel spectroscopic observations of LITTLE THINGS dwarf galaxies. *Astron. J.* (submitted).
- Glover, S. C. O. & Clark, P. C. Approximations for modelling CO chemistry in giant molecular clouds: a comparison of approaches. *Mon. Not. R. Astron. Soc.* **421**, 116–131 (2012).
- Glover, S. C. O. & Clark, P. C. Is molecular gas necessary for star formation? *Mon. Not. R. Astron. Soc.* **421**, 9–19 (2012).
- Heyer, M. *et al.* Re-examining Larson’s scaling relationships in galactic molecular clouds. *Astrophys. J.* **699**, 1092–1103 (2009).
- Billett, O. H., Hunter, D. A. & Elmegreen, B. G. Compact star clusters in nearby dwarf irregular galaxies. *Astron. J.* **123**, 1454–1475 (2002).
- Tan, J. C., Shaske, S. N. & Van Loo, S. Molecular clouds: internal properties, turbulence, star formation and feedback. *IAU Symp.* **292**, 19–28 (2013).
- Elmegreen, B. G. & Hunter, D. M. A star formation law for dwarf irregular galaxies. *Astrophys. J.* **805**, 145 (2015).
- Krumholz, M. R. Star formation in atomic gas. *Astrophys. J.* **759**, 9 (2012).
- Johnson, M. *et al.* The stellar and gas kinematics of the LITTLE THINGS dwarf irregular galaxy NGC 1569. *Astron. J.* **144**, 152 (2012).
- Turner, J. L. *et al.* Highly efficient star formation in NGC 5253 possibly from stream-fed accretion. *Nature* **519**, 331–333 (2015).
- Bekki, K. Formation of blue compact dwarf galaxies from merging and interacting gas-rich dwarfs. *Mon. Not. R. Astron. Soc.* **388**, L10–L14 (2008).
- Elmegreen, B. G., Malhotra, S. & Rhoads, J. Formation of halo globular clusters in Ly $\alpha$  emitting galaxies in the early universe. *Astrophys. J.* **757**, 9 (2012).
- Hunter, D. A. *et al.* LITTLE THINGS. *Astron. J.* **144**, 134 (2012).
- Massey, P. *et al.* A survey of local group galaxies currently forming stars. II. UVRI photometry of stars in seven dwarfs and a comparison of the entire sample. *Astron. J.* **133**, 2393–2417 (2007).
- Bolatto, A. D. *et al.* The resolved properties of extragalactic giant molecular clouds. *Astrophys. J.* **686**, 948–965 (2008).
- Wong, T. *et al.* The Magellanic Mopra Assessment (MAGMA). I. The molecular cloud population of the large magellanic cloud. *Astrophys. J.* **197**, 16 (2011).
- Solomon, P. M., Rivolo, A. R., Barret, J. & Yahil, A. Mass, luminosity, and line width relations of Galactic molecular clouds. *Astrophys. J.* **319**, 730–741 (1987).

**Acknowledgements** We wish to thank P. Massey and the Local Group Survey team for the use of their H $\alpha$  image of WLM. L. Hill made the colour composite inset in Fig. 1. M.R. would like to thank C. Herrera and J. Garcia for support with the CASA implementation to reduce the raw data and A. Rojas for support in the ALMA data reduction. M.R. is grateful to A. Leroy for providing the galaxy data to produce Fig. 3. M.R. thanks the ALMA Director for the invitation to spend her 2015 sabbatical leave at the Joint ALMA Observatory (JAO) in Santiago, where this article was finished. P.C. is grateful to L. Young and S. Madden for invaluable guidance on Herschel data reduction. M.R. wishes to acknowledge support from the Comisi  n Nacional de Investigaci  n Cient  fica y Tecnol  gica (CONICYT) through FONDECYT grant no. 1140839. M.R. is partially supported by CONICYT project BASAL PFB-06. The contributions from D.A.H. were funded by the Lowell Observatory Research Fund. P.C. acknowledges support from NASA (National Aeronautics and Space Administration) JPL RSA grant 1433776 to L. Young and grant 1456896 to D.A.H. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. NRAO is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc.

**Author Contributions** D.A.H., Principal Investigator of the ALMA proposal, identified likely CO sources from the re-processed data files using a direct search for significant emission in each frequency channel and for continuous emissions in adjacent channels. M.R. re-processed the ALMA results from the originally calibrated data delivered by ALMA to get better sensitivity and resolution, finalized the identification of emission sources, extracted spectra of the sources, produced Figs 1 and 2, and produced the measurements in Table 1. B.G.E. wrote the text of the manuscript and interpreted the main science results. E.B. oversaw the technical application of radio interferometry to molecular line mapping, and determined the noise limitations and deconvolution strategy for the angular size and velocity width measurements. J.R.C. made the size and linewidth measurements, produced the virial masses and CO luminosities, determined the main observational parameters and made Fig. 3. P.C. reduced the Herschel [CII] data and made the [CII] map used in Fig. 1. All authors contributed to the discussions leading to this manuscript.

**Author Information** This paper makes use of the following ALMA data: ADS/JAO.ALMA#2012.1.00208.S. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.G.E. (bge@us.ibm.com).

## METHODS

**ALMA observations.** We observed the  $^{12}\text{CO}(J = 1 - 0)$  transition in two regions in WLM using ALMA, located on the Chajnantor Plateau in northern Chile, during Cycle 1. Observations were carried out on 8 July 2013 and 3 April 2014. The ALMA receivers were tuned to the ground rotational transition of carbon monoxide, CO(1-0). The interferometer configuration C32-2/C32-3 provides a maximum baseline of 0.442 km. The observations were done with a spectral resolution of 122 kHz per channel ( $0.32 \text{ km s}^{-1}$ ) and total bandwidth of 468.750 MHz per baseband. The source J2258–2758 was used as a bandpass calibrator and J2357–1125 was used to calibrate amplitude and phases with time. To set the absolute flux scale, Uranus was observed. We estimated an uncertainty in absolute calibration of 10%.

The data were calibrated, mapped, and cleaned using the ALMA reduction software CASA (version 4.2.1). Rather than use the pipeline-delivered science data cubes, we redid the cleaning (that is, Fourier transform and beam deconvolution) using a better definition for masking of regions containing emission, and natural weighting to optimize sensitivity. The maximum angular scale for recovered emission was estimated to be  $15''$ .

**Identifying sources.** To make a first cut at identifying sources, we convolved the image cube to a  $1.25'' \times 1.25''$  beam and examined a wide velocity range around the velocity expected from the APEX detection. For the southeast region we expected signal around  $V_{\text{LSR}} = -120.5 \text{ km s}^{-1}$  and examined  $-130.5 \text{ km s}^{-1}$  to  $-110.5 \text{ km s}^{-1}$ . We detected candidate sources at  $-123 \text{ km s}^{-1}$  to  $-115.5 \text{ km s}^{-1}$ . For the northwest region we expected signal around  $-130.5 \text{ km s}^{-1}$  and examined  $-140.5 \text{ km s}^{-1}$  to  $-120.5 \text{ km s}^{-1}$ , detecting potential sources at  $-139 \text{ km s}^{-1}$  to  $-121.5 \text{ km s}^{-1}$ . In each velocity channel we looked for knots that had more counts than the majority of knots that were noise. Then we looked for signal in nearly the same location in successive channels, expecting coherence over at least three channels due to the Hanning smoothing that had been applied. We also generally expected the signal to build up and fade away as the channels sampled the source spectrum. With these criteria, we rated the confidence level of each candidate source as ‘confident’, ‘certain’, ‘not so certain’, or ‘uncertain’. For the southeast region, we identified nine candidate sources, six ranked as ‘confident’ or ‘certain’. In the northwest region, we identified 20 potential sources, four ranked as ‘certain’ and the rest as less certain.

On the basis of this identification, we integrated the emission in the velocity range at which CO was seen, and produced the two velocity integrated maps shown in Fig. 2 using our reduced new higher sensitivity and velocity resolution cubes. The velocity resolution of these cubes is  $0.5 \text{ km s}^{-1}$  per channel. All velocities are in the local standard of rest (LSR) system. For WLM-SE, five integrated maps were made covering a total LSR velocity range  $V_{\text{LSR}} = -121 \text{ km s}^{-1}$  to  $-115.5 \text{ km s}^{-1}$ ; the maps spanned velocities of  $-121.0 \text{ km s}^{-1}$  to  $-115.5 \text{ km s}^{-1}$ ,  $-121.5 \text{ km s}^{-1}$  to  $-119.0 \text{ km s}^{-1}$ ,  $-119.0 \text{ km s}^{-1}$  to  $-117.5 \text{ km s}^{-1}$ ,  $-118.5 \text{ km s}^{-1}$  to  $-117.0 \text{ km s}^{-1}$ , and  $-124.0 \text{ km s}^{-1}$  to  $-120.5 \text{ km s}^{-1}$ . For WLM-NW, four integrated maps were made covering  $V_{\text{LSR}} = -136.5 \text{ km s}^{-1}$  to  $-124 \text{ km s}^{-1}$ ; the individual ranges were  $-137.0 \text{ km s}^{-1}$  to  $-135.5 \text{ km s}^{-1}$ ,  $-133.5 \text{ km s}^{-1}$  to  $-130.0 \text{ km s}^{-1}$ ,  $-127.5 \text{ km s}^{-1}$  to  $-125.5 \text{ km s}^{-1}$ , and  $-127.0 \text{ km s}^{-1}$  to  $-125.5 \text{ km s}^{-1}$ . For those sources that showed emission at a 3-s.d. level or above, a spectrum was obtained integrating over an area delineated by a contour drawn at 2 s.d. (see Fig. 2) in order not to miss any genuine emission. We also produced velocity–RA and velocity–dec. maps. Inspecting the CO spectra and the velocity–position maps, we confirmed 10 CO clouds of the original 20 candidates. The remaining 10 were deemed of too low signal-to-noise to be included in this study. On each CO spectrum plot we included the H I emission FWHM velocity width and converted the H I heliocentric to LSR velocity using  $V_{\text{LSR}}$  (that is,  $V_{\text{LSR}} = V_{\text{Helioc}} - 2.5 \text{ km s}^{-1}$ ).

The total flux of the 10 clouds resolved with ALMA was compared to the CO(3-2) flux in our previous APEX observations. We converted the CO(3-2) APEX fluxes from  $\text{K km s}^{-1}$  to Jy and assumed a thermal CO(1-0)/CO(3-2) line ratio of one. For WLM-SE we recovered a similar flux of 0.42 Jy in both cases. For WLM-NW we measured an ALMA flux of 0.14 Jy while the APEX flux converted to CO(1-0) is 0.66 Jy. The difference in the northwest region can be due to a different line ratio and thus different physical conditions, or it could be from weaker emission not included in our criteria for defining CO clouds, or it could be from emission that is larger in angular extent than the largest structures measured by the interferometer and therefore absent from our maps. If we take both regions, then the measured flux with ALMA is a factor of 2 within the measured flux with APEX.

# Quadrature squeezed photons from a two-level system

Carsten H. H. Schulte<sup>1\*</sup>, Jack Hansom<sup>1\*</sup>, Alex E. Jones<sup>1</sup>, Clemens Matthiesen<sup>1</sup>, Claire Le Gall<sup>1</sup> & Mete Atatüre<sup>1</sup>

Resonance fluorescence arises from the interaction of an optical field with a two-level system, and has played a fundamental role in the development of quantum optics and its applications. Despite its conceptual simplicity, it entails a wide range of intriguing phenomena, such as the Mollow-triplet emission spectrum<sup>1</sup>, photon antibunching<sup>2</sup> and coherent photon emission<sup>3</sup>. One fundamental aspect of resonance fluorescence—squeezing in the form of reduced quantum fluctuations in the single photon stream from an atom in free space—was predicted more than 30 years ago<sup>4</sup>. However, the requirement to operate in the weak excitation regime, together with the combination of modest oscillator strength of atoms and low collection efficiencies, has continued to necessitate stringent experimental conditions for the observation of squeezing with atoms. Attempts to circumvent these issues had to sacrifice antibunching, owing to either stimulated forward scattering from atomic ensembles<sup>5,6</sup> or multi-photon transitions inside optical cavities<sup>7,8</sup>. Here, we use an artificial atom with a large optical dipole enabling 100-fold improvement of the photon detection rate over the natural atom counterpart<sup>9</sup> and reach the necessary conditions for the observation of quadrature squeezing in single resonance-fluorescence photons. By implementing phase-dependent homodyne intensity-correlation detection<sup>9–11</sup>, we demonstrate that the electric field quadrature variance of resonance fluorescence is three per cent below the fundamental limit set by vacuum fluctuations, while the photon statistics remain antibunched. The presence of squeezing and antibunching simultaneously is a fully non-classical outcome of the wave-particle duality of photons.

The minimum fluctuations in any quantum measurement of canonically conjugate variables such as position and momentum are bound by the Heisenberg uncertainty principle. Although this principle cannot be violated, the fluctuations of a single variable can be reduced below this minimum value at the expense of enhancing the fluctuations of the conjugate variable. The most widely explored realization of this non-classical phenomenon is squeezed light<sup>12</sup>, where the quadrature operators  $\hat{X}_1$  and  $\hat{X}_2$  of the electric field are the canonically conjugate operators. Relying inherently on the quadratic dependence on the bosonic creation and annihilation operators in the Hamiltonian, squeezed light is typically generated using intense lasers and macroscopic nonlinear optical media<sup>13</sup>. This form of squeezed light has multiple applications in the field of quantum optics<sup>14</sup>, one prominent example being interferometry with reduced quantum noise<sup>15,16</sup>.

In 1981, it was predicted<sup>4</sup> that the quadratic form of the Hamiltonian is not a requirement and that quadrature squeezing can also be obtained via a radically different approach: the interaction of a two-level system with a resonant light field, as described by the Jaynes–Cummings Hamiltonian. The fluctuations in one quadrature, quantified by their variance, can be reduced up to a theoretical maximum of 12.5% lower than vacuum fluctuations, while the intensity statistics remain antibunched. Unlike its nonlinear optics counterpart, this unique form of squeezed light stems from a build-up of atomic

coherence which, once mapped onto the emitted field, results in the creation of coherences between the  $n = 0$  and  $n = 1$  Fock states in the weak excitation regime. Higher number states are excluded by photon antibunching or, equivalently, by the fermionic nature of atomic operators. The simultaneous presence of antibunching and squeezing is an intriguing yet counter-intuitive effect, because single photons do not have a well-defined phase. It is the coherence with the zero-photon (vacuum) component that allows for phase-dependent effects such as the squeezing discussed here.

The two-level system we use is a voltage-controlled semiconductor quantum dot<sup>17</sup> positioned under a solid immersion lens for enhanced photon collection efficiency. Typically allowing photon detection rates well above a million photons per second, these artificial atoms obviate the immediate need for cavity coupling and consequently allow for the experimental realization of an isolated, weakly excited two-level system treated in ref. 4. With large oscillator strength, high internal quantum efficiency and negligible decoherence, semiconductor quantum dots enabled recent observations of key phenomena in quantum optics, such as antibunching<sup>18,19</sup>, formation of dressed states<sup>20–22</sup>, generation of entangled photon pairs<sup>23–25</sup> and, particularly relevant for this work, coherent single photon generation via weak excitation<sup>26,27</sup>. The strong transition dipole moment of the quantum dot compared to single atoms is also the key advantage in the detection of squeezing in resonance fluorescence. Although the maximum degree of squeezing is in principle independent of the oscillator strength of the emitter, the resultant increase in the photon scattering rate yields the necessary signal-to-noise ratio for the experimental observation of this effect (see Methods).

To generate resonance fluorescence, we excite the  $\pi^+$ -polarized neutral exciton transition of a single quantum dot resonantly at 970 nm using a frequency-stabilized tunable laser (Fig. 1a). Resonance fluorescence and the reflected laser are separated by a polarizing beam splitter and, after imparting a relative phase through a path length difference  $\Delta\ell$ , recombined via a non-polarizing beam splitter. One of the outputs of this beam splitter contains the superimposed light (SL) field

$$\hat{E}_{\text{SL}}^{(+)}(t) = \hat{E}_{\text{RF}}^{(+)}(t) + e^{i\phi} \hat{E}_{\text{LO}}^{(+)}(t) \quad (1)$$

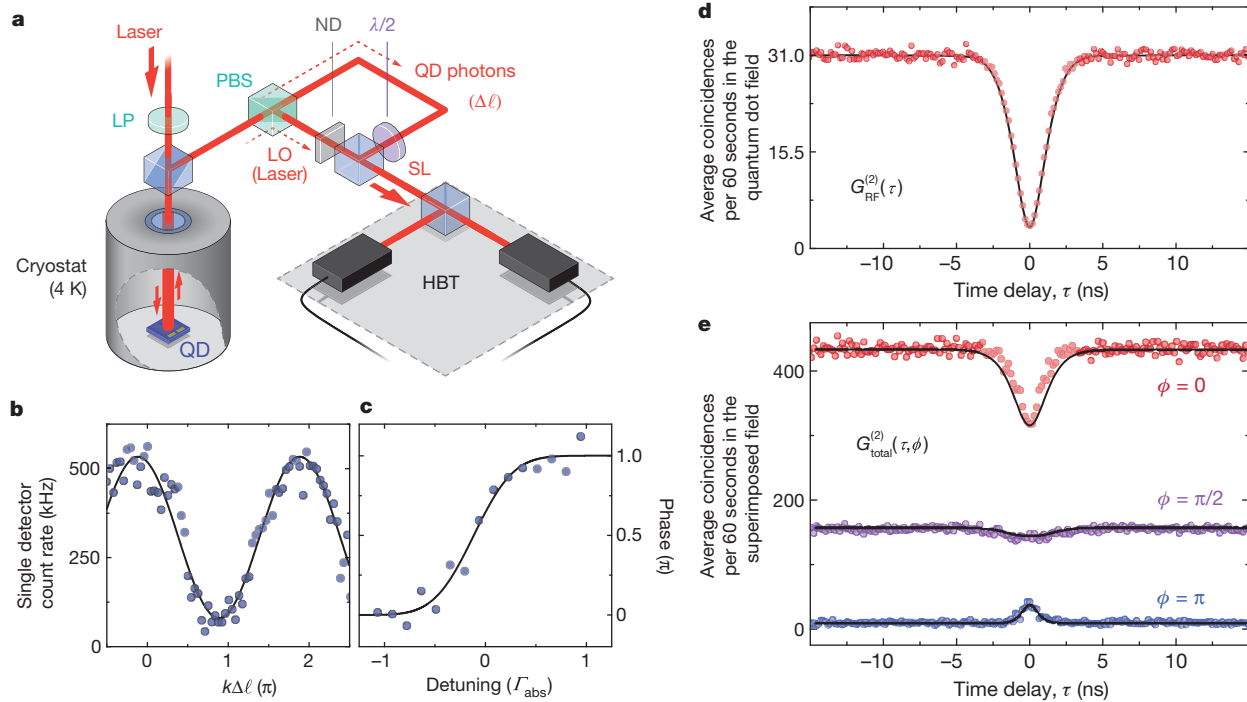
where  $t$  is time, the  $(+)$  superscripts denote the positive frequency part of the operator, the subscript LO (RF) indicates local oscillator (resonance fluorescence) and the relative amplitude and phase  $\phi$  of the local oscillator and resonance fluorescence fields can be tuned independently in the experimental scheme illustrated in Fig. 1a. Reflection and transmission coefficients and all other relative phases due to the optical setup are included in the field amplitudes and the phase  $\phi$ . Figure 1b shows the intensity measured by a single detector as a function of the interferometer-induced phase  $k\Delta\ell$ , where  $k$  is the wavenumber of the fields. The phase due to the dipolar response of the transition, which is determined by the relative detuning between the excitation laser and the transition frequency, is also included in  $\phi$ ; Fig. 1c displays the measurement of the detuning dependence of this additional phase.

The amplitude  $\langle \hat{E}(\phi) \rangle$  of a light field, where  $\phi$  is the relative phase with respect to a coherent reference field, can be represented

<sup>1</sup>Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK.

\*These authors contributed equally to this work.



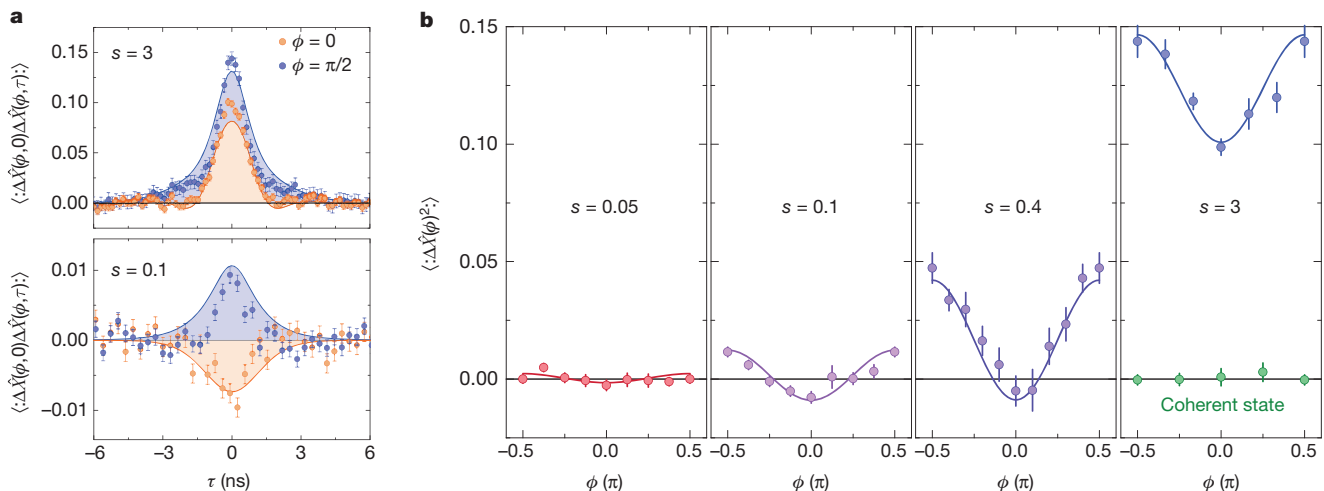


**Figure 1 | Homodyne intensity-correlation detection.** **a**, Schematic illustration of the homodyne intensity-correlation setup. QD, quantum dot; LO, local oscillator; SL, superimposed light; LP, linear polarizer; PBS, polarizing beam splitter; ND, neutral density filter;  $\lambda/2$ , half-wave plate; HBT, Hanbury Brown and Twiss correlation setup. **b**, Intensity of the superimposed light field on a single detector as a function of the interferometer phase at an excitation power of  $s = 0.1$ . At this excitation power, each detector of the Hanbury Brown and Twiss correlation setup records  $1.6 \times 10^5$  events per second from resonance fluorescence contribution alone. **c**, Dipole phase offset in interference pattern as a function of detuning between laser and quantum dot

frequency. **d**, **e**, Intensity autocorrelation measurement with the local oscillator path blocked (**d**) and unblocked (**e**). In the blocked case (**d**), the  $G_{\text{RF}}^{(2)}(\tau)$  measurement gives the expected antibunching, which is evidence of a single two-level system, regardless of  $\phi$ . The ordinate in **d** and **e** shows the coincidences in units of count rates for comparison. For the unblocked local oscillator (**e**), the interference between local oscillator and quantum dot fields leads to phase-dependent correlations, some of which contain the quadrature variance of the quantum dot field. For **b–e**, the experimental data are shown as filled circles and the theoretical simulations using a two-level master equation for the corresponding experimental conditions are shown as solid lines.

in the phase space of the conjugate variables  $\hat{X}_1$  and  $\hat{X}_2$  via  $\hat{E}(\phi) \propto \hat{X}(\phi) = (\hat{X}_1 \cos \phi + \hat{X}_2 \sin \phi)$ . These quadratures are the analogues of the dimensionless position and momentum operators and their variances, quantifying the quantum fluctuations of the electric field, and are subject to a similar uncertainty relation:  $\langle (\Delta \hat{X}_1)^2 \rangle \langle (\Delta \hat{X}_2)^2 \rangle \geq 1/16$ .

To demonstrate a squeezed quadrature ( $\Delta \hat{X}(\phi)^2 < 1/4$ ), we implement the experimental setup proposed in ref. 11, which provides a direct and convenient link between time-correlated two-photon detection and the variances of field quadratures. To detect the variances of the resonance fluorescence field quadratures  $\hat{X}_{1,2}$ , we perform



**Figure 2 | Phase-dependent quadrature variance of resonance fluorescence.** **a**, In-phase (orange) and in-quadrature (blue) normally ordered autocorrelations of the electric field quadrature fluctuations  $\langle :(\Delta \hat{X}(\phi, 0) \Delta \hat{X}(\phi, \tau)) : \rangle$  for high (top panel) and low (bottom panel) power excitation conditions. Negative values in the bottom panel verify squeezing of the in-phase electric field variance. **b**, Full phase dependence of the quadrature variances (zero time delay of the

autocorrelations), for different excitation powers. A measurement of coherent laser quadratures provides a reference for the vacuum limit of zero (green circles, right-most panel). Solid curves in all panels are theoretical simulations using a two-level master equation for the corresponding experimental conditions. Error bars represent a statistical uncertainty of one standard deviation of the correlations at long delays ( $\tau > 5$  ns).

an intensity autocorrelation on the superimposed light field  $\hat{E}_{\text{SL}}$  using a Hanbury Brown and Twiss (HBT) correlation setup. The unnormalized second-order correlation function of the superimposed light field

$$G_{\text{total}}^{(2)}(t, t + \tau) = \langle \hat{E}_{\text{SL}}^{(-)}(t) \hat{E}_{\text{SL}}^{(-)}(t + \tau) \hat{E}_{\text{SL}}^{(+)}(t + \tau) \hat{E}_{\text{SL}}^{(+)}(t) \rangle \quad (2)$$

(where  $\tau$  is the time delay and the  $(-)$  superscripts denote the negative frequency part of the operator) produces the well-known antibunched second-order correlation function of the resonance fluorescence field  $G_{\text{RF}}^{(2)}$  in the absence of a local oscillator, which demonstrates the single photon nature of resonance fluorescence<sup>2</sup>. The solid red circles in Fig. 1d display this behaviour for an excitation power of  $s = P/P_{\text{sat}} = 0.1$ , where the saturation power  $P_{\text{sat}}$  yields half of the maximum attainable resonance fluorescence intensity. The black curve is the theoretical prediction obtained with a two-level master equation and includes the detector response function, as well as sublinewidth spectral wandering of the quantum dot transition frequency<sup>28,29</sup>. In the presence of the local oscillator field,  $G_{\text{total}}^{(2)}$  displays a strong dependence on phase  $\phi$ , as shown in Fig. 1e for  $\phi = 0, \pi/2, \pi$ . While the coincidence rate at long time delays changes with  $\phi$  by more than an order of magnitude, the correlation behaviour at short time delays evolves from a dip to a peak as a function of  $\phi$ .

The total correlation function  $G_{\text{total}}^{(2)}$  contains five terms with  $|E_{\text{LO}}|^n$  for  $n = 0, 1, 2, 3, 4$ . We separate their contributions via their unique dependence on the time delay ( $\tau$ ) and relative phase ( $\phi$ ), as well as via direct measurement of the zeroth-order contribution (Fig. 1d); see Supplementary Information. The second-order contribution is directly proportional to the normally ordered autocorrelation of  $\Delta\hat{X}(\phi, t)$ :

$$\Delta G_2^{(2)}(\tau) \propto \langle : \Delta\hat{X}(\phi, 0) \Delta\hat{X}(\phi, \tau) : \rangle \quad (3)$$

where  $\langle : \dots : \rangle$  denotes the normal ordering of the field operators. The zero-delay value of equation (3) hence yields the normally ordered variance  $\langle : (\Delta\hat{X}(\phi))^2 : \rangle$  of the electric field quadrature<sup>11</sup>. This variance is zero for vacuum and coherent fields and the existence of quadrature squeezing is manifested in a negative-valued normally ordered variance:  $\langle : (\Delta\hat{X}(\phi))^2 : \rangle < 0$ .

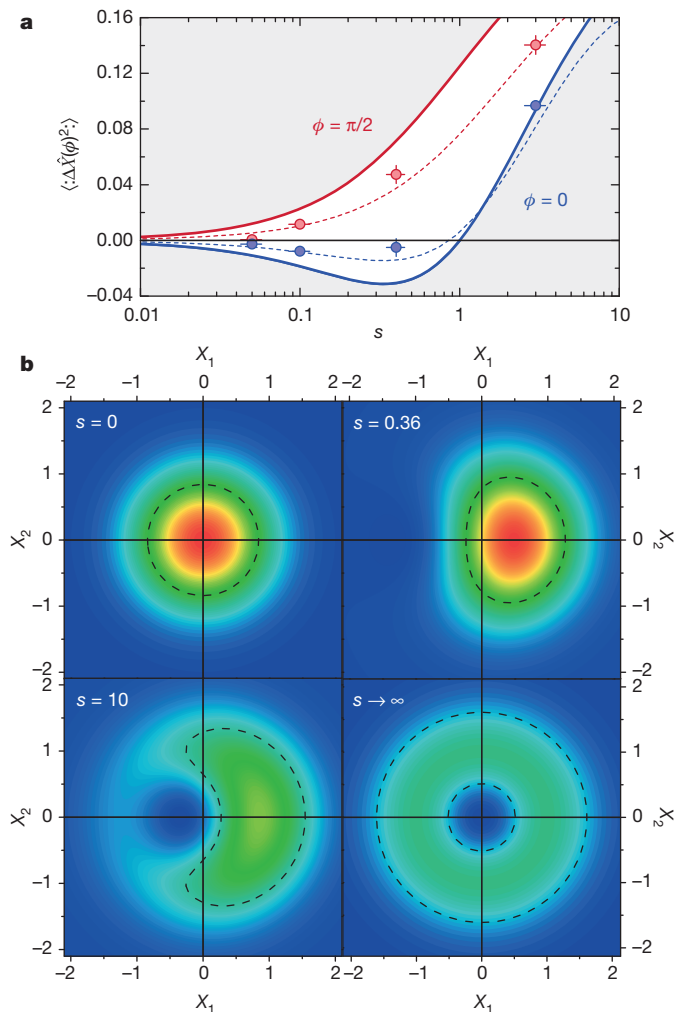
Figure 2a presents the autocorrelation of the in-phase ( $\phi = 0$ ) and in-quadrature ( $\phi = \pi/2$ ) fluctuations for high excitation power ( $s = 3$ , upper panel of Fig. 2a). The normally ordered variance ( $\tau = 0$ ) of resonance fluorescence in the high-power regime is positive-valued regardless of the phase. This indicates that the quantum fluctuations are enhanced above the vacuum level, as expected. In stark contrast, the low-power regime ( $s = 0.1$ , lower panel of Fig. 2a) yields negative values for the normally ordered variance for  $\phi = 0$ . This reduction of quantum fluctuations below the vacuum limit is the verification of quadrature squeezing in this measurement. As dictated by the Heisenberg uncertainty relations, this squeezing is accompanied by increased fluctuations, that is, antisqueezing, in the other quadrature. Both features decay on a timescale of the order of the excited state lifetime<sup>30</sup>.

Figure 2b shows how the normally ordered variance evolves as a function of  $\phi$ , for excitation powers ranging from  $s = 0.05$  (left-most panel) to  $s = 3$  (right-most panel). A measurement with a weak laser of similar intensity is displayed in the rightmost plot as green circles. As expected for a coherent state, this measurement yields a normally ordered variance of zero, independent of  $\phi$ . The squeezing vanishes<sup>4</sup> for  $s \geq 1$ , yielding larger fluctuations than vacuum for any  $\phi$ . Although we observe the phase dependence of the quadrature variance at all excitation powers, the window of opportunity for measuring negative values is restricted to a very small  $\phi$  range in the low-excitation regime ( $s < 1$ ), which highlights the challenges that have been associated with the observation of squeezing in resonance fluorescence since its prediction.

Figure 3a summarizes the power dependence of the normally ordered quadrature variance extrema. The solid blue (red) curve

represents the theoretically predicted behaviour of the in-phase (in-quadrature) field variance for an ideal two-level system. The maximum possible squeezing is limited to 12.5% (0.58 dB) below vacuum fluctuations at  $s = 0.36$ . The dashed curves depict how the ideal two-level system behaviour is modified, owing to the combined effects of the finite timing resolution of our detection system and the phase uncertainty of our interferometer (see Methods). All variance extrema we measured are commensurate with these predictions, which confirms that deviation from the solid curve is predominantly due to limitations of a technical nature. The maximum degree of squeezing we measure is  $3.1\% \pm 1\%$  (0.14 dB) below vacuum noise at an excitation power of  $s = 0.1$ . This value corresponds to 40% of the theoretically obtainable limit set by the blue solid curve at this excitation power.

The transformation of the state of light with excitation power is best visualized by the calculated Wigner functions presented in Fig. 3b. The



**Figure 3 | Excitation-power dependence of quadrature squeezing.** **a**, The measured (symbols) and the theoretical (solid curves) normally ordered quadrature variances as a function of the excitation power  $s$  for  $\phi = 0$  and  $\phi = \pi/2$ . The dashed curves illustrate the effect of phase noise and finite timing resolution in our experiment on an otherwise ideal two-level system (see Supplementary Information). Horizontal error bars, uncertainty in the excitation power; vertical error bars, s.d. of the correlations, as in Fig. 2. **b**, Wigner functions for different excitation powers. The transition from vacuum state ( $s = 0$ ) to a mixture of vacuum and single photon Fock state ( $s \rightarrow \infty$ ) displays non-symmetric forms at intermediate excitation regimes ( $s = 0.36$  and  $s = 10$ ). For  $s = 0.36$ , the spread of the Wigner function along the quadrature variable  $X_1$  is clearly less than that of vacuum, a manifestation of quadrature squeezing. The dashed lines depict the contour at 50% of the maximum value at each power.

spread of these phase-space distributions along a given polar angle  $\phi$  is indicative of the variance of the corresponding field quadrature  $\hat{X}(\phi) = \hat{X}_1 \cos(\phi) + \hat{X}_2 \sin(\phi)$ . The Wigner function for the vacuum state (top left) shows a symmetric form with no phase dependence. At intermediate powers (top right and bottom left), the symmetry breaks down and a  $\phi$ -dependence arises in the spread of the Wigner function, linked to the generation of atomic coherence (see Supplementary Information). This phase dependence, in combination with the antibunched nature of resonance fluorescence, leads to a reduced variance of the electric field for a phase angle of  $\phi = 0$  (as seen in, for example, the Wigner function at  $s = 0.36$ ). In the high-power regime ( $s \rightarrow \infty$ ), the field becomes a statistical mixture of the  $n = 0$  and  $n = 1$  Fock states and the steady-state phase dependence is lost completely.

We have shown that resonance fluorescence from a two-level system can comprise a single photon stream with below-vacuum quantum fluctuations of the field. While this appears counter-intuitive, owing to the impossibility of associating a well-defined phase to single photons, the probabilistic nature of coherent photon scattering in the weak excitation regime allows the emitted photons to be in a coherent superposition of Fock states  $|0\rangle$  and  $|1\rangle$ . The emergence of phase correlations in this regime endows resonance fluorescence with the coexistence of photon antibunching and quadrature squeezing. Our simultaneous observation of these two phenomena can therefore be interpreted as a quantum mechanical manifestation of the complementary particle and wave natures of light, respectively, with no classical analogues.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 1 April; accepted 23 June 2015.**

**Published online 31 August 2015.**

- Schuda, F., Stroud, C. R. Jr & Hercher, M. Observation of the resonant Stark effect at optical frequencies. *J. Phys. B* **7**, L198–L202 (1974).
- Kimble, H. J., Dagenais, M. & Mandel, L. Photon antibunching in resonance fluorescence. *Phys. Rev. Lett.* **39**, 691–695 (1977).
- Höffges, J. T., Baldauf, H. W., Lange, W. & Walther, H. Heterodyne measurement of the resonance fluorescence of a single ion. *J. Mod. Opt.* **44**, 1999–2010 (1997).
- Walls, D. F. & Zoller, P. Reduced quantum fluctuations in resonance fluorescence. *Phys. Rev. Lett.* **47**, 709–711 (1981).
- Heidmann, A. & Reynaud, S. Squeezing in the many atom resonance fluorescence emitted in the forward direction: application to photon noise reduction. *J. Phys. (Paris)* **46**, 1937–1948 (1985).
- Lu, Z. H., Bali, S. & Thomas, J. E. Observation of squeezing in the phase-dependent fluorescence spectra of two-level atoms. *Phys. Rev. Lett.* **81**, 3635–3638 (1998).
- Raizen, M. G., Orozco, L. A., Xiao, M., Boyd, T. L. & Kimble, H. J. Squeezed-state generation by the normal modes of a coupled system. *Phys. Rev. Lett.* **59**, 198–201 (1987).
- Ourjoumtsev, A. *et al.* Observation of squeezed light from one atom excited with two photons. *Nature* **474**, 623–626 (2011).
- Gerber, S. *et al.* Intensity-field correlation of single-atom resonance fluorescence. *Phys. Rev. Lett.* **102**, 183601 (2009).
- Ou, Z. Y., Hong, C. K. & Mandel, L. Detection of squeezed states by cross correlation. *Phys. Rev. A* **36**, 192–196 (1987).
- Vogel, W. Homodyne correlation measurements with weak local oscillators. *Phys. Rev. A* **51**, 4160–4171 (1995).
- Yuen, H. P. Two-photon coherent states of the radiation field. *Phys. Rev. A* **13**, 2226–2243 (1976).
- Teich, M. C. & Saleh, B. E. A. Squeezed states of light. *Quantum Opt.* **1**, 153–199 (1989).
- Walls, D. F. Squeezed states of light. *Nature* **306**, 141–146 (1983).
- Caves, C. M. Quantum limits on noise in linear amplifiers. *Phys. Rev. D* **26**, 1817–1839 (1982).
- Goda, K. *et al.* A quantum-enhanced prototype gravitational-wave detector. *Nature Phys.* **4**, 472–476 (2008).
- Warburton, R. J. *et al.* Giant permanent dipole moments of excitons in semiconductor nanostructures. *Phys. Rev. B* **65**, 113303 (2002).
- Michler, P. *et al.* Quantum correlation among photons from a single quantum dot at room temperature. *Nature* **406**, 968–970 (2000).
- Kim, J., Benson, O., Kan, H. & Yamamoto, Y. A single-photon turnstile device. *Nature* **397**, 500–503 (1999).
- Xu, X. *et al.* Coherent optical spectroscopy of a strongly driven quantum dot. *Science* **317**, 929–932 (2007).
- Vamivakas, A. N., Zhao, Y., Lu, C.-Y. & Atatüre, M. Spin-resolved quantum-dot resonance fluorescence. *Nature Phys.* **5**, 198–202 (2009).
- Flagg, E. B. *et al.* Resonantly driven coherent oscillations in a solid-state quantum emitter. *Nature Phys.* **5**, 203–207 (2009).
- Akopian, N. *et al.* Entangled photon pairs from semiconductor quantum dots. *Phys. Rev. Lett.* **96**, 130501 (2006).
- Young, R. J. *et al.* Improved fidelity of triggered entangled photons from single quantum dots. *New J. Phys.* **8**, 29 (2006).
- Müller, M., Bounouar, S., Jöns, K. D., Glässl, M. & Michler, P. On-demand generation of indistinguishable polarization-entangled photon pairs. *Nature Photon.* **8**, 224–228 (2014).
- Matthiesen, C., Vamivakas, A. N. & Atatüre, M. Subnatural linewidth single photons from a quantum dot. *Phys. Rev. Lett.* **108**, 093602 (2012).
- Matthiesen, C. *et al.* Phase-locked indistinguishable photons with synthesized waveforms from a solid-state source. *Nature Commun.* **4**, 1600 (2013).
- Kuhlmann, A. V. *et al.* Charge noise and spin noise in a semiconductor quantum device. *Nature Phys.* **9**, 570–575 (2013).
- Stanley, M. J. *et al.* Dynamics of a mesoscopic nuclear spin ensemble interacting with an optically driven electron spin. *Phys. Rev. B* **90**, 195305 (2014).
- Loudon, R. Squeezing in resonance fluorescence. *Opt. Commun.* **49**, 24–28 (1984).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge financial support from the University of Cambridge, the European Research Council ERC Consolidator Grant Agreement No. 617985 and the EU-FP7 Marie Curie Initial Training Network S3NANO. C.M. acknowledges Clare College Cambridge for financial support through a Junior Research Fellowship. We thank E. Clarke, M. Hugues and the EPSRC National Centre for III-V Technologies for the wafer and C. Baune, R. Moghadas Nia, W. Vogel, G. Rempe, H. J. Carmichael and A. Ourjoumtsev for discussions.

**Author Contributions** C.H.H.S. and M.A. devised the experiment, C.H.H.S., J.H., A.E.J., C.M. and C.L.G. performed the experiments, C.H.H.S., J.H. and C.L.G. developed the models and analysed the data, all authors contributed to the discussion of the results and the manuscript preparation. C.H.H.S. and C.M. processed the quantum dot device.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A. (ma424@cam.ac.uk).



## METHODS

**Interferometer.** A frequency- and power-stabilized single-mode laser is used to resonantly excite the neutral exciton transition of the quantum dot. The emitted photons are collected in a confocal dark-field microscope, where the laser is separated from the quantum dot emission by means of two crossed polarizers. The second polarizer is implemented as a polarizing beam splitter, which enables the use of the attenuated laser field as local oscillator. The light field in the quantum dot arm of the interferometer consists typically of  $<1\%$  laser photons and  $>99\%$  resonance fluorescence photons; we therefore neglect the laser background in the quantum dot photon mode. Likewise, any quantum dot photons in the local oscillator output can be neglected because the excitation laser intensity before attenuation is several orders of magnitude larger than the resonance fluorescence intensity. In the fringe measurement in Fig. 1a, the spatial path difference  $\Delta\ell \approx 11$  cm is kept constant while laser and quantum dot frequency are tuned continuously to change the interferometer phase. This form of phase control is enabled by the tunability of the quantum dot resonance via the DC Stark effect<sup>17</sup> and increases the long-term stability of the interferometer, which contains no moving parts. The visibility of the interferometer for high-power laser light amounts to near unity, but is reduced in Fig. 1b to 73.8%, owing to incoherent photon emission as well as an inadvertent mismatch of resonance fluorescence and local oscillator intensities. Additionally, the visibility is reduced for low count rates<sup>9</sup>, which makes the use of a bright single photon source and high photon collection efficiencies crucial for our experiments. The collection efficiency of our optical setup is 1%. This value is calculated from the obtained count rates and the 0.58-ns excited-state lifetime of the quantum dot used in this work.

**Post-selection.** The intensities of resonance fluorescence and local oscillator fields are kept equal in all of our measurements. To ensure the absence of laser photons in the quantum dot mode, the laser background is measured once a minute for 2 s. To this end, the local oscillator path is blocked and the quantum dot, which is embedded in a Schottky diode structure, is tuned off resonance via the quantum confined Stark effect (QCSE)<sup>17</sup>. Furthermore, the intensity of the quantum dot emission is monitored continuously during the measurements to detect spectral wandering of the quantum dot transition<sup>28,29</sup>. This is done by filtering out the phonon sideband (PSB) and detecting it on a third single photon detector. The measured correlation histograms ( $G_{\text{total}}^{(2)}(\tau)$ ) are saved once a minute. We perform a

post-selection of histograms with a threshold on the mean PSB count rate and another threshold on the measured laser leakage in the quantum dot arm. In the experiments shown in Fig. 2, the laser is kept on resonance but the relative phase is not actively controlled in the interferometer. Instead, phase-dependent measurements are performed by using the individual detector intensities as a measure of the interferometer phase and relying on the wandering of the phase due to temperature drifts on timescales of typically  $\gtrsim 30 \text{ min } \pi^{-1}$ . To bin the data, we perform a reference measurement of the interference fringes by scanning the laser frequency while keeping the quantum dot on resonance using the QCSE. An example measurement of the interference fringes obtained in this way may be seen in Fig. 1b. To have equal sized phase bins, we use intensity bins of varying size that are proportional to the derivative of  $\cos^2(\phi/2)$ . This phase binning is sign-invariant, that is, it cannot distinguish between positive or negative phases and therefore bins data into values between 0 and  $+\pi$ . This does not affect our measurement because all correlation functions are symmetric in phase around 0. The data points shown at negative phases in Fig. 2b are measured between  $\phi = 0.5\pi$  and  $\phi = \pi$ , and have been shifted by  $-\pi$ .

**Reduction of measured degree of squeezing.** Although the conditioning nature of the measurement should render it robust against low detection efficiencies, this is not true in practice, and several effects reduce the measured degree of squeezing compared to the theoretical limit for resonance fluorescence. Low photon numbers in the interferometer reduce the fringe visibility, but this only affects the signal-to-noise ratio in our measurements. However, low count rates also lead to higher shot noise which can increase the error in the phase binning protocol. This leads to a decrease in the detected degree of squeezing by introducing mixing between quadratures. Other sources of phase noise include spectral wandering of the quantum dot transition leading to fluctuations in the dipole phase, and any interferometric instability on timescales shorter than the histogram saving time. Finite timing resolution of the correlation setup also leads to a decreased visibility of the features at zero time delay, and further reduces our measured value for squeezing. We have independently measured the timing resolution of the Hanbury Brown and Twiss correlation setup with a mode-locked pulsed laser source ( $<3$ -ps pulse width) for different mean count rates. The extent of phase noise from different sources is harder to quantify and is used as a fitting parameter in the theoretical curves in Fig. 2.

# The most incompressible metal osmium at static pressures above 750 gigapascals

L. Dubrovinsky<sup>1\*</sup>, N. Dubrovinskaia<sup>2\*</sup>, E. Bykova<sup>1,2</sup>, M. Bykov<sup>2</sup>, V. Prakapenka<sup>3</sup>, C. Prescher<sup>3</sup>, K. Glazyrin<sup>4</sup>, H.-P. Liermann<sup>4</sup>, M. Hanfland<sup>5</sup>, M. Ekholm<sup>6,7</sup>, Q. Feng<sup>6,7</sup>, L. V. Pourovskii<sup>6,8</sup>, M. I. Katsnelson<sup>9,10</sup>, J. M. Wills<sup>11</sup> & I. A. Abrikosov<sup>7,12</sup>

Metallic osmium (Os) is one of the most exceptional elemental materials, having, at ambient pressure, the highest known density and one of the highest cohesive energies and melting temperatures<sup>1</sup>. It is also very incompressible<sup>2–4</sup>, but its high-pressure behaviour is not well understood because it has been studied<sup>2–6</sup> so far only at pressures below 75 gigapascals. Here we report powder X-ray diffraction measurements on Os at multi-megabar pressures using both conventional and double-stage diamond anvil cells<sup>7</sup>, with accurate pressure determination ensured by first obtaining self-consistent equations of state of gold, platinum, and tungsten in static experiments up to 500 gigapascals. These measurements allow us to show that Os retains its hexagonal close-packed structure upon compression to over 770 gigapascals. But although its molar volume monotonically decreases with pressure, the unit cell parameter ratio of Os exhibits anomalies at approximately 150 gigapascals and 440 gigapascals. Dynamical mean-field theory calculations suggest that the former anomaly is a signature of the topological change of the Fermi surface for valence electrons. However, the anomaly at 440 gigapascals might be related to an electronic transition associated with pressure-induced interactions between core electrons. The ability to affect the core electrons under static high-pressure experimental conditions, even for incompressible metals such as Os, opens up opportunities to search for new states of matter under extreme compression.

The platinoid *5d* transition elements Re, Os, and Ir are the densest and stiffest metals<sup>1,7</sup>. Although a short-lived claim<sup>4</sup> that Os is stiffer than diamond<sup>8</sup> was subsequently disproven<sup>2</sup>, there remains scientific interest in the high-pressure behaviour of Os: the bulk modulus values measured by different groups vary substantially (395–435 GPa)<sup>2,3,5,6,9</sup>, and controversy surrounds reports of a possible pressure-induced isostructural Lifshitz<sup>10</sup> transition (also called an electronic topological transition, ETT) in Os. The ETT arises when distortion of the electronic band structure by an external perturbation results in a topological modification of the Fermi surface.

Os has a hexagonal close-packed (hcp) structure, with two unit cell parameters (*a* and *c*) fully defining the atomic arrangement at a given pressure. The observation<sup>3</sup> of an anomaly in the compressibility and pressure dependence of the *c/a* ratio around 25 GPa was attributed to an ETT, but subsequent experimental studies<sup>2,5,6</sup> at pressures to about 60 GPa failed to detect anomalies and found<sup>6,11</sup> instead that texturing or non-hydrostatic conditions can greatly complicate the interpretation of X-ray diffraction data; indeed, experimental artefacts may mimic isostructural transitions. Theoretical studies of Os have so far also resulted in an inconsistent picture of its high-pressure beha-

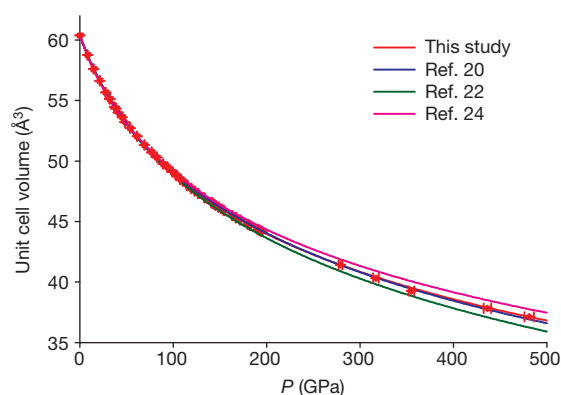
viour. Reports of possible single or multiple anomalies in *c/a* ratio at pressures ranging from 9 GPa to over 25 GPa have been both attributed<sup>3</sup> and not attributed<sup>12</sup> to ETT; one study<sup>13</sup> found evidence for multiple ETTs at pressures up to 130 GPa that did not affect the compressional behaviour of Os, and a further study<sup>14</sup> concluded that there are no peculiarities in the pressure-driven evolution of the atomic and electronic structure of Os. The inconsistencies in the theoretical high-pressure behaviour of Os mirror the difficulties encountered when probing the high-pressure behaviour of *3d* hcp metals such as Zn and  $\epsilon$ -Fe (refs 15–19). These difficulties add further interest to detailed studies of the *5d* element Os, to allow a broader comparison of the crystal chemistry of hcp metals.

Although Os has been experimentally studied<sup>3,4,6</sup> at pressures up to about 75 GPa, this pressure range is far too narrow to explore the behaviour of one of the most incompressible metals. Data collected at multi-megabar pressures are highly desired, and can be obtained using the double-stage diamond anvil cell (dsDAC) technique<sup>7</sup>, which generates the required ultra-high static pressures. Under such conditions, pressure determination is based on the equation of state (EOS) of one or several standards mixed with the sample being studied (see Methods). The absolute accuracy of the EOS measurements—which is particularly important when aiming to compare experiment against theory—cannot be higher than the accuracy of the static pressure scale. Shock-wave and ramp compression experiments<sup>20,21</sup> achieve terapascal pressures, but only at high temperatures. The EOSs obtained in experiments below 100 GPa, using different dynamic and static methods and for different standards, tend to agree to within 2–3 GPa, but discrepancies increase with pressure and frequently reach unacceptable levels of the order of tens of gigapascals at a pressure around 0.5 TPa (Fig. 1, Extended Data Fig. 1, examples in ref. 22). Much more accurate results are obtained when using standards with internally consistent EOSs, especially if the materials used as pressure markers have different (or, even better, contrasting) elastic properties<sup>22–24</sup>, as do gold, platinum, and tungsten.

We conducted experiments in conventional and dsDACs on Au–Pt mixtures at pressures up to 500 GPa, and on Au–W and Pt–W mixtures up to approximately 200 GPa (Fig. 1, Extended Data Figs 1, 2). With Au as the pressure marker<sup>20</sup>, we fitted the pressure–volume (*P*–*V*) data of Pt using the third-order Birch–Murnaghan (Fig. 1) and the Vinet EOSs (Extended Data Table 1). Both EOSs provide equally good fits, yielding parameters for Pt very close to those obtained from shock-wave data<sup>20</sup> (Fig. 1, Extended Data Table 1). Pt and Au were used as pressure markers in two independent powder X-ray diffraction experiments to study the EOS of tungsten (Extended

<sup>1</sup>Bavarian Research Institute of Experimental Geochemistry and Geophysics, University of Bayreuth, D-95440 Bayreuth, Germany. <sup>2</sup>Laboratory of Crystallography, University of Bayreuth, D-95440 Bayreuth, Germany. <sup>3</sup>Center for Advanced Radiation Sources, University of Chicago, Illinois 60437 Argonne, USA. <sup>4</sup>Photon Sciences, Deutsches Elektronen-Synchrotron (DESY), D-22603 Hamburg, Germany. <sup>5</sup>European Synchrotron Radiation Facility, BP 220, Grenoble F-38043, France. <sup>6</sup>Swedish e-Science Research Centre (SeRC), Linköping University, SE-58183 Linköping, Sweden. <sup>7</sup>Department of Physics, Chemistry and Biology (IFM), Linköping University, SE-58183 Linköping, Sweden. <sup>8</sup>Centre de Physique Théorique, CNRS, École Polytechnique, 91128 Palaiseau, France. <sup>9</sup>Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525AJ Nijmegen, The Netherlands. <sup>10</sup>Department of Theoretical Physics and Applied Mathematics, Ural Federal University, Mira street 19, Ekaterinburg, 620002, Russia. <sup>11</sup>Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA. <sup>12</sup>Materials Modeling and Development Laboratory, National University of Science and Technology 'MISIS', 119049 Moscow, Russia.

\*These authors contributed equally to this work.

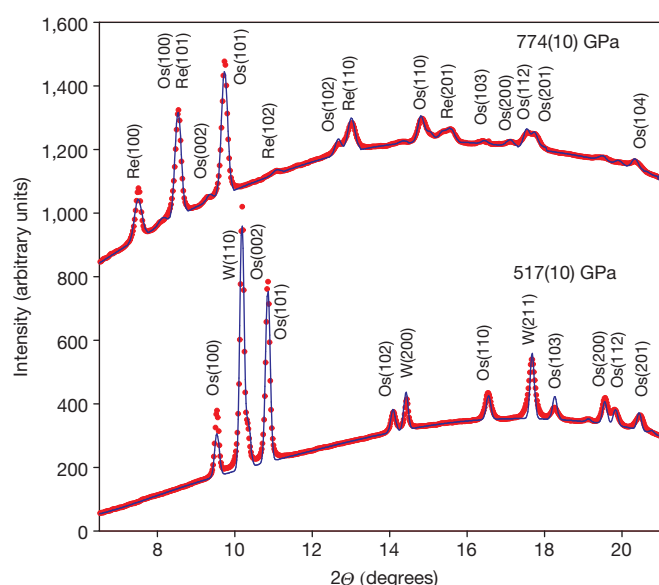


**Figure 1 | The dependence of the unit cell volume of Pt on pressure  $P$ .**

Experimental data points (different symbols correspond to different runs, Au was used as a pressure marker in all experiments, error bars show standard deviations) were fitted using the third-order Birch–Murnaghan EOS (red line, volume at ambient temperature and pressure  $V_0 = 60.389(3) \text{ Å}^3$  per unit cell, EOS parameters  $K_{300} = 274(2) \text{ GPa}$ ,  $K' = 5.23(3)$ ). The blue, green, and pink lines are from refs 20, 22, and 24, respectively.

Data Figs 1, 2 and Extended Data Table 1). These experiments yield  $P$ – $V$  data that are in remarkable agreement with each other and provide an accurate EOS of W for pressures up to 200 GPa (Extended Data Fig. 1 and Extended Data Table 1). Our data are in close agreement with the EOS of W obtained using a thermodynamic assessment<sup>23</sup>, but disagree with extrapolations of the EOSs above about 100 GPa derived from experiments below 1 Mbar (Extended Data Fig. 1). These results confirm the necessity of EOS calibration well above 100 GPa for ultra-high-pressure studies and provide (Extended Data Table 1) a self-consistent set of EOSs for Au, Pt, and W that we use in our experiments with Os.

We studied the compressional behaviour of Os using powder X-ray diffraction experiments in conventional DACs at pressures up to about 200 GPa that were conducted in Ne (Au and W pressure markers, Extended Data Fig. 2) or He (Pt pressure marker). Similar experiments in a dsDAC reached a pressure of 774(10) GPa (according to the Os EOS; see below), where the number in parentheses is one standard deviation; Os adopted the hcp structure even at this pressure (Fig. 2). The only successful experiments in dsDACs with a pressure marker were using the Os–W mixture (Fig. 2), where we achieved a pressure slightly above 500 GPa. The data collected in different runs are consistent (Fig. 3), and the dependence of the unit cell volume on pressure does not show any obvious anomaly and is well described by the third-order Birch–Murnaghan (Fig. 3a) or Vinet EOSs (Extended Data Table 1). The bulk modulus (399(6) GPa fitted with the third-order Birch–Murnaghan EOS) is in agreement with the results of refs 2 (Fig. 3a) and 3 (Extended Data Fig. 3). We do, however, observe two anomalies at approximately 150 GPa and 440 GPa in the ratio of the lattice parameters,  $c/a$ , when compressing Os (Fig. 3b). Although these are at the detection limit of our experimental set-up, the anomaly at about 150 GPa was reproduced in three independent runs, and the one above 400 GPa in two runs. Fitting of the  $P$ – $V$  data by the third-order Birch–Murnaghan EOS within three pressure ranges—below 120 GPa, between 170 GPa and 400 GPa, and above 400 GPa—gives an interesting result: although fitting over the first and second pressure ranges produces similar EOS parameters ( $K_{300} = 397(3) \text{ GPa}$ ,  $K' = 4.07(4)$  and  $K_{300} = 416(5) \text{ GPa}$ ,  $K' = 3.8(1)$ , respectively, where the subscript indicates room temperature and the prime denotes differentiation with respect to pressure), we obtain different parameters at pressures above 400 GPa ( $K_{300} = 293(5) \text{ GPa}$ ,  $K' = 5.4(1)$ ). These experimentally observed peculiarities are not artefacts and require an explanation: substantial changes of compressibility with pressure are quite



**Figure 2 | Diffraction patterns of the samples compressed in dsDACs.**

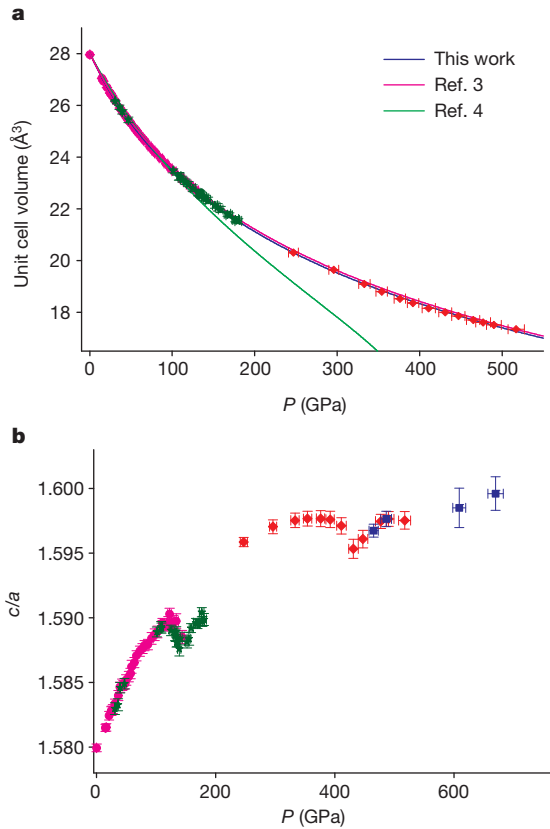
Lower curve, mixture of Os ( $a = 2.3232(3) \text{ Å}$ ,  $c = 3.7099(11) \text{ Å}$ ) and W ( $a = 2.6649(2) \text{ Å}$ , pressure marker); X-ray wavelength of  $0.3344 \text{ Å}$ . Upper curve, Os ( $a = 2.2488(2) \text{ Å}$ ,  $c = 3.555(2) \text{ Å}$ , pressure determined from our Os EOS); X-ray wavelength of  $0.2898 \text{ Å}$ . Reflections of highly textured Re at approximately 110 GPa are due to the Re support of the secondary anvil. Experimental data are shown by red dots; continuous blue curves are simulations using the full-profile (GSAS) software.  $\theta$  is the diffraction angle; the labels on the peaks indicate the Miller indices of the corresponding diffraction reflection of the given metal; diffraction lines of secondary anvils are not visible; the pressures given above the curves designate at which pressure in a dsDAC the diffraction patterns were collected.

common for materials undergoing pressure-induced spin crossover<sup>19</sup>, but are not expected for heavy  $5d$  elements such as Os.

The behaviour of the  $c/a$  ratio of Os suggests that an ETT is the most likely explanation for the observed peculiarities<sup>19</sup>. In first-principles electronic structure calculations for Os at different compressions in the framework of dynamical mean-field theory<sup>25</sup> (see Methods), we observe two ETTs at  $\Gamma$  and L points of the hcp Brillouin zone at pressures of approximately 100 GPa and 180 GPa, respectively (Extended Data Figs 4, 5). This result agrees with the pressure range in which the first  $c/a$  ratio anomaly appears experimentally (Fig. 3b). The behaviour of the bands at  $\Gamma$  and L points for increasing pressure is remarkably similar to that observed recently in the isoelectronic hcp Fe (ref. 19). We observe a rather strong influence of many-electron effects beyond the local density approximation within density functional theory (DFT) on the band structure at  $\Gamma$  and L points even though Os can be classified as a weakly correlated metal (Extended Data Fig. 4). The inclusion of correlations between  $5d$  electrons moves the ETTs at these high symmetry points to higher pressure (Extended Data Fig. 5), improving the agreement with experiment. However, increasing the pressure further does not lead to any new ETTs up to a pressure  $P = 477 \text{ GPa}$  (Extended Data Fig. 6), even when including the spin–orbit interaction (Methods and Extended Data Figs 7, 8). Consequently, the origin of the second anomaly in the  $c/a$  ratio seen in Fig. 3b remains unknown.

In solid metals, the outermost (valence) electrons are no longer associated with their respective atoms and instead form electronic bands that bond the atoms together. Because the inner-core electrons remain tightly bound to their nuclei and do not contribute to the bonds, they are often considered irrelevant when determining the properties of the material. But compression increases the overlap between the electronic clouds, as seen in the plot of the electronic density of states (DOS) of Os in Fig. 4, where the low-lying localized  $5p$  and  $4f$  states start to interact with each other at  $P = 392 \text{ GPa}$ . This





**Figure 3 | Experimental dependence of the unit cell volume and the ratio of the lattice parameters  $c/a$  of Os on pressure  $P$ .** a, b, Different symbols correspond to different pressure markers in different runs (solid magenta dots, Au and W pressure markers; green stars, Pt marker; red diamonds, W pressure marker; blue square, Os itself; error bars show standard deviations). Experimental data points were fitted using the third-order Birch–Murnaghan EOS (blue line,  $V_0 = 28.02(4) \text{ Å}^3$  per unit cell,  $K_{300} = 399(6) \text{ GPa}$ ,  $K' = 4.04(4)$ ). The magenta and green lines are from refs 3 and 4, respectively.

pressure for the core-level crossing (CLC) transition is in good agreement with the pressure at which we see the  $c/a$  ratio anomaly in our experiments, suggesting that interactions between  $5p$  and  $4f$  states might cause the observed peculiarity.

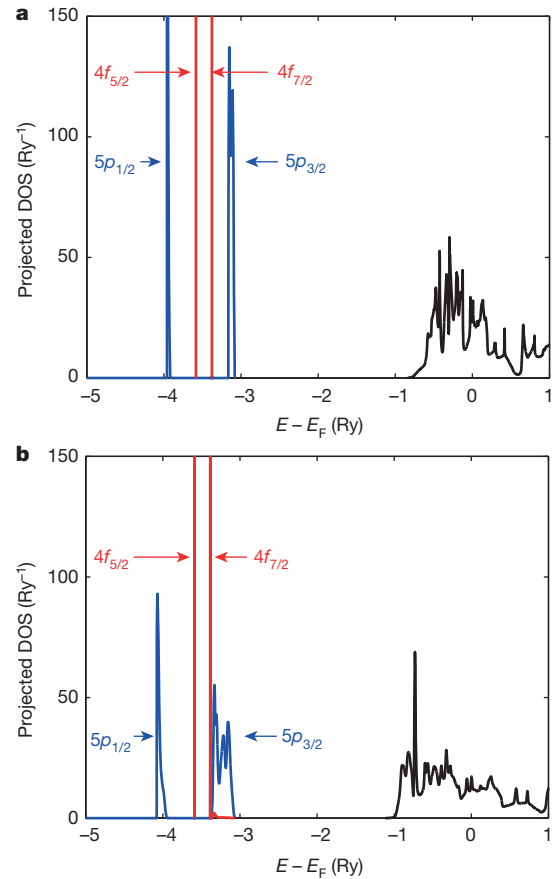
The effect of the CLC transition on thermodynamic properties must be indirect. Using the pseudopotential transformation<sup>26</sup>, one can establish a relation between the smooth part of the valence orbitals  $\tilde{\psi}_i^v$ , which determine the bonding, and the core orbitals  $\psi_j^c$ . The former are solutions of the Schrödinger-like equation

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + \hat{V} \right] \tilde{\psi}_i^v(\mathbf{r}) = \varepsilon_i^v \tilde{\psi}_i^v(\mathbf{r}) \quad (1)$$

where  $\hbar$  is the reduced Planck constant,  $m$  is the mass of the electron,  $\mathbf{r}$  is the position vector,  $\varepsilon_i^v$  are the eigenvalues of the orbitals, and  $\hat{V} = V + \hat{V}^R$  is the pseudopotential, with  $V$  an effective potential that acts on the electron. The non-local operator  $\hat{V}^R$  acts upon  $\tilde{\psi}_i^v$  according to

$$\hat{V}^R \tilde{\psi}_i^v(\mathbf{r}) = \sum_j (\varepsilon_i^v - \varepsilon_j^c) \langle \psi_j^c | \tilde{\psi}_i^v \rangle \psi_j^c(\mathbf{r}) \quad (2)$$

Well-separated and atomic-like core states are often considered ‘frozen’, with no pressure-dependent influence on the valence states; this explains the success of modern pseudopotential approaches when studying matter under extreme conditions. But substantial reconstruction of inner  $5p$  and  $4f$  states can affect chemical bonding and thereby structural properties of solids via modification of the non-local operator  $\hat{V}^R$  in equation (2).



**Figure 4 | Calculated electronic density of states (DOS) of Os as a function of energy  $E$  (relative to the Fermi energy  $E_F$ ).** a, b, DOS at pressures of 0 GPa (a) and 392 GPa (b). Experimental lattice parameters were used in the calculations.  $6s$ ,  $6p$ ,  $5d$ , and  $5f$  electrons form well-defined bands near the Fermi energy at all the pressures examined in this study.  $5p$  electrons occupy  $5p_{1/2}$  and  $5p_{3/2}$  states, which are split owing to spin–orbit interaction;  $4f$  electrons occupy  $4f_{5/2}$  and  $4f_{7/2}$  states. They behave as core electrons forming fully localized states at  $P = 0 \text{ GPa}$ . However, at  $P = 392 \text{ GPa}$ , the  $5p$  states are broadened. Importantly, one clearly sees in b that  $5p_{3/2}$  and  $4f_{7/2}$  states start to interact with each other, and the CLC transition takes place. This interaction might be responsible for the peculiarity observed in our experiments at ultra-high pressure.

Even in the simplest model of polarized ionic cores of small radius, where the pseudopotential reduces to a local operator, its Fourier component  $v_{ps}(q)$  is<sup>27</sup>

$$v_{ps}(q) = Z v_C(q) - \sum_{\mathbf{k}} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\alpha(i\omega)}{(1 + \omega_p^2/\omega^2)^2} v_C(k) \times v_C(|\mathbf{k} - \mathbf{q}|) \frac{k(k - q)}{-i\omega + \varepsilon_{k-q}} \quad (3)$$

where  $Z$  is the ionic charge,  $\mathbf{q}$  is the wave vector of the Fourier component of the pseudopotential,  $v_C(q) = 4\pi e^2/q^2$  is the Fourier component of the Coulomb interaction,  $\mathbf{k}$  is the electron wave vector,  $\varepsilon_k$  is the dispersion relation for the conduction electrons,  $\omega_p$  is their plasma frequency, and  $\alpha(i\omega)$  is the ion core polarizability at imaginary frequencies. In this model, reconstruction of the inner states affects the electron–ion interaction via the function  $\alpha(i\omega)$ ; Os under ultra-high pressure seems to provide the first real example of this kind of effect. In a sense this effect resembles the well-known phenomenon of atomic collapse<sup>28</sup>, where, as one moves across the periodic table,  $f$  electron states move from their outer effective potential well to the inner well (which explains the existence of rare-earth and actinide groups).

Although the effects we document are much weaker, the unavoidable strong mixing of  $4f$  and  $5p$  states at the CLC changes atomic polarizability and therefore the potential of the ion–electron interaction. Figure 4b shows that the overlap of  $5p$  states increases with pressure, which equates to an increase of their effective radius. This should make their contribution to the ionic polarizability larger, and translate the effect of the CLC—a transition that is expected to be common, at least for the transition metals of the sixth period—to the valence electrons.

Relating electronic transitions to lattice parameter anomalies is generally nontrivial (Methods). For instance, a conventional ETT caused by a Fermi surface topology change need not result in a visible anomaly in the pressure dependence of lattice parameters at temperature  $T = 0$  (Extended Data Figs 9, 10, Extended Data Table 1) because the ETT gives rise to a pronounced anomaly in the third derivative of the thermodynamic potential and induces some kinks in the second derivative (this is the reason it is often called “the second-and-a-half”-order transition<sup>10</sup>). But the ETT gives rise to a divergence in the thermal expansion, resulting in an observable anomaly in the  $c/a$  ratio at finite temperatures (Methods). A similar situation may exist regarding the new type of electronic transition observed in our experiments at ultrahigh pressure, even though the effect of the crossing of deep  $4f$  and  $5p$  levels should be weaker than the effect arising from the Fermi surface topology modification; the experimentally observed anomaly in the  $c/a$  ratio at approximately 440 GPa is weaker than the one at approximately 150 GPa (Fig. 3b).

Our findings demonstrate that extreme compression can change the nature of core electrons. This effect has been examined in static high-pressure experiments for soft simple metals like Li (ref. 29) and Na (ref. 30), but recent ramp compression experiments on diamond<sup>21</sup> have indicated that qualitatively new static pressure levels are needed to affect the core electrons of less compressible transition metals or covalently bonded materials. By compressing Os, one of most incompressible metals, to over 770 GPa, we were able to access this regime and observe a new type of electronic transition, the CLC transition, that involves pressure-induced interactions between core electrons, and leads to observable changes of the material properties. We believe that the ability to reach sufficiently high pressure levels to affect the core electrons of transition metals in static high-pressure experiments will open up opportunities in the search for new states of matter.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 January; accepted 17 June 2015.

Published online 24 August 2015.

1. Young, D. A. *Phase Diagrams of the Elements* (Univ. of California Press, 1991).
2. Kenichi, T. Bulk modulus of osmium: high-pressure powder x-ray diffraction experiments under quasihydrostatic conditions. *Phys. Rev. B* **70**, 012101 (2004).
3. Occelli, F. *et al.* Experimental evidence for a high-pressure isostructural phase transition in osmium. *Phys. Rev. Lett.* **93**, 095502 (2004).
4. Cynn, H., Klepeis, J. E., Yoo, C.-S. & Young, D. A. Osmium has the lowest experimentally determined compressibility. *Phys. Rev. Lett.* **88**, 135701 (2002).
5. Armentrout, M. M. & Kavner, A. Incompressibility of osmium metal at ultrahigh pressures and temperatures. *J. Appl. Phys.* **107**, 093528 (2010).
6. Godwal, B. K., Yan, J., Clark, S. M. & Jeanloz, R. High-pressure behavior of osmium: an analog for iron in Earth's core. *J. Appl. Phys.* **111**, 112608 (2012).
7. Dubrovinsky, L., Dubrovinskaya, N., Prakapenka, V. & Abakumov, A. Implementation of micro-ball nanodiamond anvils for high-pressure studies above 6 Mbar. *Nature Commun.* **3**, 1163 (2012).
8. Occelli, F., Loubeyre, P. & Letoullec, R. Properties of diamond under hydrostatic pressures up to 140 GPa. *Nature Mater.* **2**, 151–154 (2003).
9. Voronin, G. A., Pantea, C., Zerda, T. W., Wang, L. & Zhao, Y. Thermal equation-of-state of osmium: a synchrotron X-ray diffraction study. *J. Phys. Chem. Solids* **66**, 706–710 (2005).

10. Lifshitz, I. M. *Zh. Eksp. Teor. Fiz.* **38**, 1569 (1960); Anomalies of electron characteristics of a metal in the high-pressure region. *Sov. Phys. JETP* **11**, 1130 (1960).
11. Weinberger, M. B., Tolbert, S. H. & Kavner, A. Osmium metal studied under high pressure and nonhydrostatic stress. *Phys. Rev. Lett.* **100**, 045506 (2008).
12. Ma, Y. M. *et al.* Electronic and crystal structures of osmium under high pressure. *Phys. Rev. B* **72**, 174103 (2005).
13. Koudela, D., Richter, M., Möbius, A., Koepernik, K. & Eschrig, H. Lifshitz transitions and elastic properties of osmium under pressure. *Phys. Rev. B* **74**, 214103 (2006).
14. Liang, Y. & Fang, Z. First-principles study of osmium under high pressure. *J. Phys. Condens. Matter* **18**, 8749–8759 (2006).
15. Kenichi, T. Structural study of Zn and Cd to ultrahigh pressures. *Phys. Rev. B* **56**, 5170–5179 (1997).
16. Kenichi, T. Absence of the  $c/a$  anomaly in Zn under high pressure with a helium-pressure medium. *Phys. Rev. B* **60**, 6171–6174 (1999).
17. Klotz, S., Braden, M. & Besson, J. M. Is there an electronic topological transition in zinc under high pressure? *Phys. Rev. Lett.* **81**, 1239–1242 (1998).
18. Steinle-Neumann, G., Stixrude, L. & Cohen, R. Absence of lattice strain anomalies at the electronic topological transition in zinc at high pressure. *Phys. Rev. B* **63**, 054103 (2001).
19. Glazyrin, K. *et al.* Importance of correlation effects in hcp iron revealed by a pressure-induced electronic topological transition. *Phys. Rev. Lett.* **110**, 117206 (2013).
20. Yokoo, M. *et al.* Ultrahigh-pressure scales for gold and platinum at pressures up to 550 GPa. *Phys. Rev. B* **80**, 104114 (2009).
21. Smith, R. F. *et al.* Ramp compression of diamond to five terapascals. *Nature* **511**, 330–333 (2014).
22. Dorogokupets, P. I. & Dewaele, A. Equations of state of MgO, Au, Pt, NaCl-B1, and NaCl-B2: internally consistent high-temperature pressure scales. *High Press. Res.* **27**, 431–446 (2007).
23. Sokolova, T. S., Dorogokupets, P. I. & Litasov, K. D. Self-consistent pressure scales based on the equations of state for ruby, diamond, MgO, B2–NaCl, as well as Au, Pt, and other metals to 4 Mbar and 3000 K. *Russ. Geol. Geophys.* **54**, 181–199 (2013).
24. Holmes, N. C., Moriarty, J. A., Gathers, G. R. & Nellis, W. J. The equation of state of platinum to 660 GPa. *J. Appl. Phys.* **66**, 2962–2967 (1989).
25. Kotliar, G. *et al.* Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.* **78**, 865–951 (2006).
26. Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* 210 (Cambridge Univ. Press, 2004).
27. Katsnelson, M. I. & Trefilov, A. V. The effect of van der Waals and Born-Mayer interactions of ionic cores on the atomic properties of the alkali metals. *Fiz. Tverdogo Tela* **30**, 3299–3310 (1988).
28. Karaziy, R. I. The collapse of the excited electron orbit and the peculiarities of the atomic spectra. *Usp. Fiziol. Nauk* **135**, 79–115 (1981).
29. Matsuoka, T. & Shimizu, K. Direct observation of a pressure-induced metal-to-semiconductor transition in lithium. *Nature* **458**, 186–189 (2009).
30. Ma, Y. *et al.* Transparent dense sodium. *Nature* **458**, 182–185 (2009).

**Acknowledgements** L.D. and N.D. acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) and the Federal Ministry of Education and Research (BMBF), Germany. N.D. thanks the DFG for funding through the Heisenberg Program and the DFG project number DU 954-8/1, and the BMBF for grant number 5K13WC3 (Verbundprojekt 05K2013, Teilprojekt 2, PT-DESY). M.E., Q.F., and I.A.A. acknowledge support from the Swedish Foundation for Strategic Research programme SRL grant numbers 10-0026, the Swedish Research Council (VR) grant numbers 621-2011-4426, the Swedish Government Strategic Research Area Grant Swedish e-Science Research Centre (SeRC), and in Materials Science “Advanced Functional Materials” (AFM). The work was supported by the Ministry of Education and Science of the Russian Federation (grant number 14.Y26.31.0005). The simulations were carried out using supercomputer resources provided by the Swedish national infrastructure for computing (SNIC). M.I.K. acknowledges financial support from the ERC Advanced grant number 338957 FEMTO/NANO and from NWO via a Spinoza Prize. Portions of this work were performed at GeoSoilEnviroCARS (Sector 13), Advanced Photon Source (APS), Argonne National Laboratory. GeoSoilEnviroCARS is supported by the National Science Foundation - Earth Sciences (EAR-1128799) and Department of Energy - GeoSciences (DE-FG02-94ER14466). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract number DE-AC02-06CH11357.

**Author Contributions** L.D., N.D., E.B., M.B., V.P., C.P., K.G., H.-P.L., and M. H. conducted the experiments. L.D. and N.D. processed the experimental data. M.E., Q.F., L.V.P., M.I.K., J.M.W., and I.A.A. performed the theoretical analysis. The manuscript was written by L.D. and I.A.A. with contributions from all other authors. All the authors commented on drafts and have approved the final version of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.D. (Leonid.Dubrovinsky@uni-bayreuth.de) or I.A.A. (Igor.Abrikosov@ifm.liu.se).

## METHODS

**Materials.** Au (99.9995% purity, 1- $\mu$ m particle size), Pt (99.999% purity, 2- $\mu$ m particle size), W (99.999% purity, 1- $\mu$ m particle size), and Os (99.99% purity, 1- $\mu$ m particle size) powders were purchased from Goodfellow Inc.

**Diamond anvil cell experiments.** For all our experiments we used piston-cylinder BX90 type DACs with a large optical aperture produced at BGI. Diamonds with culet sizes of 250  $\mu$ m or 300  $\mu$ m were used in conventional DAC experiments below 70 GPa or as primary anvils in double-stage DACs. Compression of Au, Pt, W, and Os powders in a He or Ne pressure medium above 100 GPa was conducted with bevelled diamonds with 120- $\mu$ m culets. The secondary anvils were produced by direct conversion of glassy carbon balls with diameters of 10–20  $\mu$ m. Re or Ir were used as gasket materials. Gaskets were indented to a thickness of 20–35  $\mu$ m in different experiments and holes with diameters of 125–150  $\mu$ m were drilled into the centre of the indentation. As a pressure transmitting medium, He and Ne was loaded at pressures of 1.2–1.4 kbar; in some experiments with double-stage DACs, liquid paraffin wax was used. Cavities in diamond anvils for experiments with double-stage DACs were made by a picosecond pulsed laser.

**X-ray diffraction measurements.** We conducted *in situ* X-ray high-pressure experiments at the Bayerisches Geoinstitut (Germany), at ID09 at ESRF (France), at ECB at PETRA III (Germany), and at 13-IDD at Advanced Photon Source, APS (USA). At the Bayerisches Geoinstitut, we obtained powder X-ray diffraction data with a system consisting of a Rigaku FRD high-brilliance generator (90 kW) and APEX CCD Area Detector. The MoK $\alpha$  radiation (tube voltage 60 kV, tube current 55 mA, cathode gun 0.1  $\times$  0.1 mm $^2$ ) was focused with MaxFlux X-ray optics and further collimated down to a FWHM beam size of about 30  $\mu$ m. At the ID09 at ESRF, the data were collected with the MAR555 detector using an X-ray beam with a wavelength of approximately 0.41 Å and a beam size down to 5  $\times$  5  $\mu$ m $^2$ . At the ECB at PETRA III, the data were collected with the Perkin Elmer detector using an X-ray beam with a wavelength of approximately 0.29 Å and a beam size as small as 1.5  $\times$  1.5  $\mu$ m $^2$ . At the 13-IDD station (GSECARS), we used a MAR-165 CCD area detector and a highly focused beam (about 3  $\times$  4  $\mu$ m $^2$ ) with a wavelength of 0.3344 Å. The collected images were integrated using the Fit2D or GADDS programs to obtain a conventional diffraction pattern. Data analysis was conducted using the GSAS package.

**Computational details.** To include electronic correlation effects for the partially-filled Os 5d band beyond the standard DFT framework we used an local density approximation + dynamical mean-field theory (LDA+DMFT) approach<sup>25,31</sup>. This approach is based on a full-potential linear augmented plane-wave + local orbitals technique as implemented in the Wien2k code<sup>40</sup> in conjunction with the DMFT implementation provided by the TRIQS package<sup>31–34</sup>. Our LDA+DMFT framework is fully self-consistent in the charge density. The LDA+DMFT calculations were performed within the scalar-relativistic approximation and using a  $\mathbf{k}$ -mesh with 32  $\times$  32  $\times$  32 points in the full Brillouin zone. The spin–orbit coupling was not included because LDA calculations show that it has a negligible effect on the electronic structure in the vicinity of the Fermi level. The DMFT quantum impurity problem was solved using the numerically-exact imaginary-time hybridization-expansion continuous-time quantum Monte Carlo (CT-QMC) method<sup>35</sup>. A large number of Monte Carlo cycles, more than 512 million, were performed to obtain a well converged DMFT local self-energy. We adopted a stochastic version of the maximum entropy method<sup>36</sup> for the analytical continuation of the CT-QMC self-energy to the real frequency axis. For the Coulomb interaction strength  $U$  and Hund's coupling constant  $J$  we used the values  $U = 2.8$  eV and  $J = 0.55$  eV that are estimated in ref. 37. The qualitative results of our LDA+DMFT calculations are not very sensitive to the exact values of  $U$  and  $J$ . We used the 'around mean-field form'<sup>38</sup> for the double counting correction, which is suitable for weakly correlated metallic systems.

In calculations of band structure at the level of DFT<sup>39</sup> within the LDA or semi-local generalized gradient approximation (GGA), we used two complementary methods, the full potential (linear) augmented plane waves + local orbitals method as implemented in the Wien2k code<sup>40</sup> and the electronic-structure method<sup>41</sup> RSPt. Both are all-electron methods, which do not impose any approximations on the shape of the one-electron potential, and they are known to generate very similar results. The former method allows us to directly compare the LDA and LDA+DMFT results. These methods are particularly suited to high-pressure calculations because the basis functions for any energy, including nominally deep core states, can be treated as 'valence' states.

For calculations with the Wien2k code, we used a  $\mathbf{k}$ -mesh consisting of 32  $\times$  32  $\times$  32  $\mathbf{k}$ -points in the full Brillouin zone. The size of the plane-wave basis set is given by the cutoff parameter  $K_{\text{max}}$ . In our calculations, we kept the product between  $K_{\text{max}}$  and the radius of the real-space muffin-tin spheres to  $K_{\text{max}} \times R_{\text{MT}} = 10$ . At pressures of 0 GPa, 134 GPa, 247 GPa, and 477 GPa, we set  $R_{\text{MT}} = 2.5$  atomic units (a.u.), 2.34 a.u., 2.27 a.u., and 2.16 a.u., respectively.

Wien2k LDA results give the most direct comparison to our LDA+DMFT results, because same computational scheme at the LDA level is used.

To calculate the band-projected density of states (DOS) we used RSPt<sup>41</sup>. RSPt is an all-electron, full-potential electronic structure method that uses a basis of site-centred spherical waves (a generalization of augmented muffin-tin orbitals) in the self-consistent Kohn–Sham formalism<sup>40</sup> to calculate the electron density and total energy. We used LDA, as well as two gradient corrected functionals, AM05<sup>42,43</sup> and PBE<sup>44</sup>—well-known examples of the accuracy DFT can achieve<sup>45</sup>. We used RSPt to predict the electron DOS of Os in the hexagonal close-packed (hcp) structure at pressures up to about 700 GPa. At these pressures, we find that a basis corresponding to 4f, 5s, 5p, 5d, 5f, 6s, and 6p atomic states is sufficiently complete; adding 4d functions to the valence yields negligible changes in the calculated DOS. Bases are scalar-relativistic: the spin–orbit interaction is included variationally.

Calculations of band structure and DOS were carried out at the experimental values of the lattice parameters for each value of pressure. For a few lattice parameters, the calculations were carried out using both the Wien2k code with LDA, and the RSPt code with GGA-PBE; the obtained DOS are very similar. The electronic structure calculated for a fixed lattice parameter is known to be quite insensitive to the use of LDA or GGA in the calculations<sup>46</sup>. When calculating the  $c/a$  ratio using RSPt, enthalpy was optimized at fixed pressure on a grid of pressures. Using Wien2k, energy was minimized at fixed volume.

**Influence of correlation effects.** To underline the importance of correlation effects in calculations of the electronic structure of Os, we compare the results of LDA+DMFT and LDA calculations in Extended Data Figs 4 and 5. The difference is quite noticeable. In our LDA calculations, we do not observe any ETTs at the  $\Gamma$  point as pressure increases, because at 0 GPa the corresponding band is already well above the Fermi level. This result is in contrast to that of ref. 13, where an ETT was found at this point, but in agreement with ref. 14, in which no ETT was reported. This discrepancy is due to the different values for the lattice constant that were used in the LDA calculations. We used experimental room-temperature values of the lattice parameters. Using the LDA lattice constants from ref. 13, we recover the band energy below the Fermi level at the  $\Gamma$  point. Assuming the GGA lattice constants from ref. 14, we reproduce the results of this work at the  $\Gamma$  point.

Our LDA calculations also predict that the L-point ETT occurs at a much smaller pressure than our LDA+DMFT calculations, around 100 GPa (Extended Data Figs 4 and 5). In ref. 14, this band at the L point was predicted to be just below the Fermi level at 129 GPa; no ETTs are reported in this pressure range. Using the same lattice constants as in ref. 14, we reproduce these results within LDA. We also find that along the L–H line, the band energy is just above the Fermi level, but this part of the Brillouin zone is not shown in ref. 14. Thus, we attribute discrepancies between the LDA- and GGA-based studies to differences in the EOS rather than to the exchange–correlation potential.

The discussion above shows that the electronic structure at the  $\Gamma$  and L points is quite sensitive to volume changes, and that the occurrence, as well as the position of ETTs in the LDA/GGA picture, depends sensitively on the accuracy of the assumed EOS. The accuracy of the calculated EOS in Os depends on the approximation for the electron–electron interactions used in calculations, as is discussed below. In view of this uncertainty, the most reliable description of the electronic structure is obtained using the experimentally measured lattice parameters<sup>46</sup>. We adopted this strategy, and show all the electronic structure plots at the experimental lattice parameters. We did not detect any substantial difference between the electronic structure calculated with the Wien2k code and with RSPt methods.

**Influence of relativistic effects.** Because Os is a heavy element, the importance of the spin–orbit coupling (SOC) should be investigated. Using LDA, we calculated the band structure in both the scalar-relativistic approximation and with the inclusion of SOC using the Wien2k code<sup>40</sup>; the results are shown in Extended Data Figs 7 and 8. Some of the bands are split as a result of the inclusion of SOC. However, no new features are seen in the immediate vicinity of the Fermi level. In both cases, we find that one ETT has already taken place at the L point of the hcp Brillouin zone at a pressure of 134 GPa (Extended Data Fig. 7). We do not see any new ETTs upon increasing the pressure to 477 GPa (Extended Data Fig. 8). Instead, we see that the agreement between LDA and LDA+DMFT improves at high pressure, as expected, because the importance of correlation effects decreases with increasing pressure. This observed agreement indicates the internal consistency of our calculations.

**Calculated equation of state.** The calculations of the EOS and the lattice parameters using the LDA+DMFT approach are very time consuming, and their numerical accuracy is insufficient to distinguish weak peculiarities of the lattice parameters, such as the  $c/a$  ratio<sup>47</sup>. We therefore focus on the results obtained within the LDA and GGA of the DFT, and compare our results with experiment, as well as with data available in the literature (see Extended Data Fig. 9 and 10 and Extended Data Table 1).



**Relationship between electronic transitions and anomalies in lattice parameters.** Let us first consider an ETT due to the change of the Fermi surface topology. Although we have shown the importance of many-electron effects for hcp Os, they mainly influence band positions at the  $\Gamma$  and L points, while the metal remains weakly correlated. Thus, we can use the one-electron picture for a qualitative discussion. For three-dimensional systems the main effect of correlations on the ETT is the change of coefficients at the singularities<sup>48</sup>. The character of the anomalies due to the ETT is different at low temperatures (lower than typical phonon energies) and at high ones. The initial anomaly is in the DOS at the Fermi energy, which within the one-electron picture is a square root singularity, for example,  $\delta N(E_F) \propto \sqrt{z}\theta(z)$ , where  $z$  is the distance between the Fermi energy  $E_F$  and the Van Hove singularity  $E_C$  ( $z = E_F - E_C$ ), and  $\theta(z)$  is the Heaviside function<sup>10</sup>. In the case of an appearance of a new hole pocket<sup>19</sup> below the critical volume  $V_{ETT}$ , the change in the DOS is  $\delta N(E_F) \approx (V_{ETT} - V)^{1/2}$ . The anomaly yields a sharp peculiarity in the third derivative of the thermodynamic potential  $\Omega$ , and induces some kinks in the second derivative. However, it does not necessarily lead to a visible peculiarity of the pressure dependence of the lattice parameters at  $T = 0$  K, in agreement with our calculations (Extended Data Fig. 10). Still, in hcp metals the effect of the ETT on the lattice parameters can be detected experimentally at finite temperature, owing to the anisotropy of the thermal expansion coefficients  $\alpha_c$  and  $\alpha_a$  along the  $c$  and  $a$  directions of the crystal lattice, respectively.

Indeed,  $\alpha_c$  and  $\alpha_a$  can be evaluated from the phonon  $F^{\text{phon}}$  and the electron  $F^{\text{el}}$  contributions to the free energy of the hcp metal<sup>49,50</sup>:

$$\alpha_a = \frac{1}{3V(B_{11}B_{22} - B_{12}^2)} \left[ -(B_{22} + B_{12}) \frac{\partial^2 F^*}{\partial T \partial \epsilon_v} + (B_{12} + B_{11}) \frac{\partial^2 F^*}{\partial T \partial \epsilon_c} \right] \quad (4)$$

$$\alpha_c = \frac{1}{3V(B_{11}B_{22} - B_{12}^2)} \left[ -(B_{22} - 2B_{12}) \frac{\partial^2 F^*}{\partial T \partial \epsilon_v} + (B_{12} - 2B_{11}) \frac{\partial^2 F^*}{\partial T \partial \epsilon_c} \right]$$

where  $F^* = F^{\text{phon}} + F^{\text{el}}$ ,  $T$  is the temperature,  $V$  is the volume,  $\epsilon_v \equiv \ln(V)$ ,  $\epsilon_c \equiv \ln(c/a)$ , and the coefficients  $B_{ij}$  are defined via the hcp elastic constants  $C_{ij}$  according to

$$B_{11} = \frac{2}{9} \left( C_{11} + C_{12} + \frac{1}{2} C_{33} + 2C_{13} \right)$$

$$B_{22} = \frac{2}{9} (C_{11} + C_{12} + 2C_{33} - 4C_{13})$$

$$B_{12} = \frac{1}{9} (C_{33} + C_{13} - C_{11} - C_{12})$$

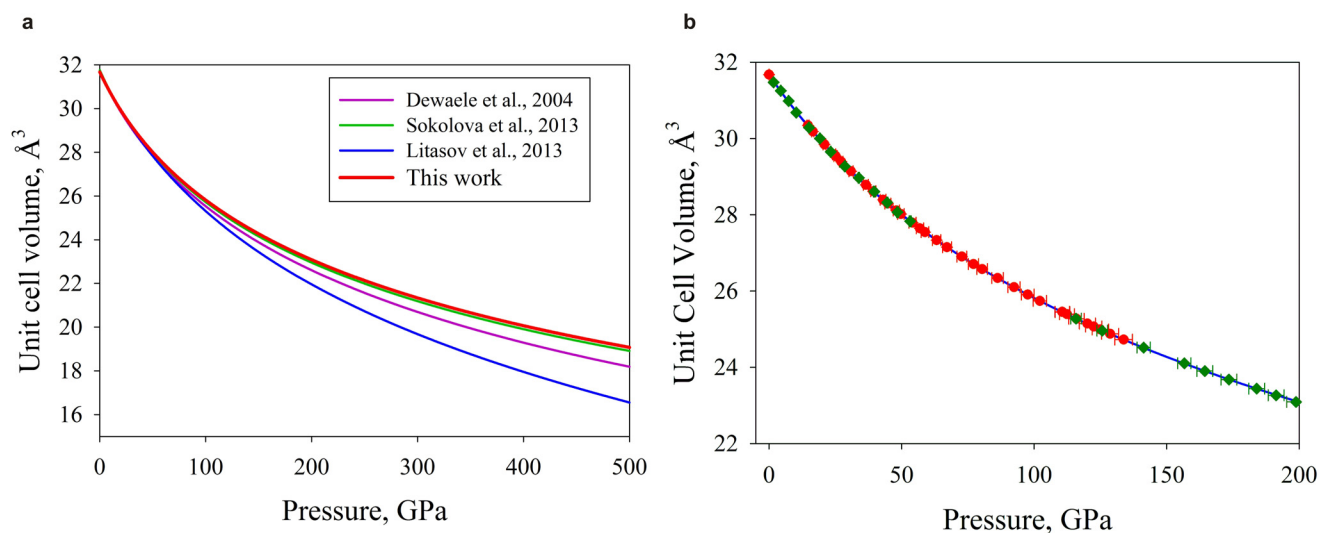
Thus, the anisotropy of the thermal expansion is an inherent property of an hcp metal, but it should become stronger at the point of the ETT. The contribution due to the ETT to the electronic Grüneisen parameter and thus to the electronic thermal expansion coefficient is proportional to  $\partial \delta N(E_F) / \partial u \propto 1/\sqrt{z}$  (where  $u$  is the dilatation) and is divergent at the point of the ETT. However, this contribution is dominant only at a very low temperatures (typically of the order of 10 K or below), and phonon contribution should be considered. For acoustic phonons with small wave vectors, the Grüneisen parameter and thus the thermal expansion coefficient are proportional to  $\partial C_{ij} / \partial u$  (typically there is only one shear modulus that is especially sensitive to the ETT, and it is the modulus that should be considered<sup>51</sup>). Because<sup>51</sup>  $\partial C_{ij} \propto \delta N(E_F)$ , the low-temperature phonon thermal expansion coefficient is as singular as the electronic one. The anomaly in phonon frequencies all over the Brillouin zone, except at the vicinity of the  $\Gamma$  point, is determined by the change of effective interatomic potential via the screening anomaly, which is weaker and proportional to  $(-z)^{3/2}\theta(-z)$  (ref. 48); as a result, at high enough temperatures, we expect a square-root singularity in the thermal expansion coefficient.

The temperature at which a crossover occurs from a stronger to a weaker singularity depends on peculiarities of the phonon spectra of a particular material. The peculiarity of the  $c/a$  ratio appears in the experiment as an effect that is integrated over the temperature, and therefore the contribution from strong peculiarities at low temperature should remain visible in room-temperature experiments. For example, although for hcp Ti the crossover temperature is about 150 K<sup>50</sup>, the temperature induced change of the  $c/a$  ratio between 0 K and room temperature is of the order of 0.5%, judging from the experiment reported<sup>49</sup>. In hcp Fe, the behaviour of the bands at the  $\Gamma$  and L points is very similar to the case of Os and the peculiarity of  $c/a$  is also visible<sup>19</sup>. The difference between hcp Fe and Os is that in the former metal, the bands cross the Fermi level almost at the same pressure, whereas in the latter, the crossover takes place within a larger pressure range.

A CLC transition discovered in this work may also lead to an observable peculiarity of the  $c/a$  ratio. The anisotropy of the thermal expansion in hcp metals is

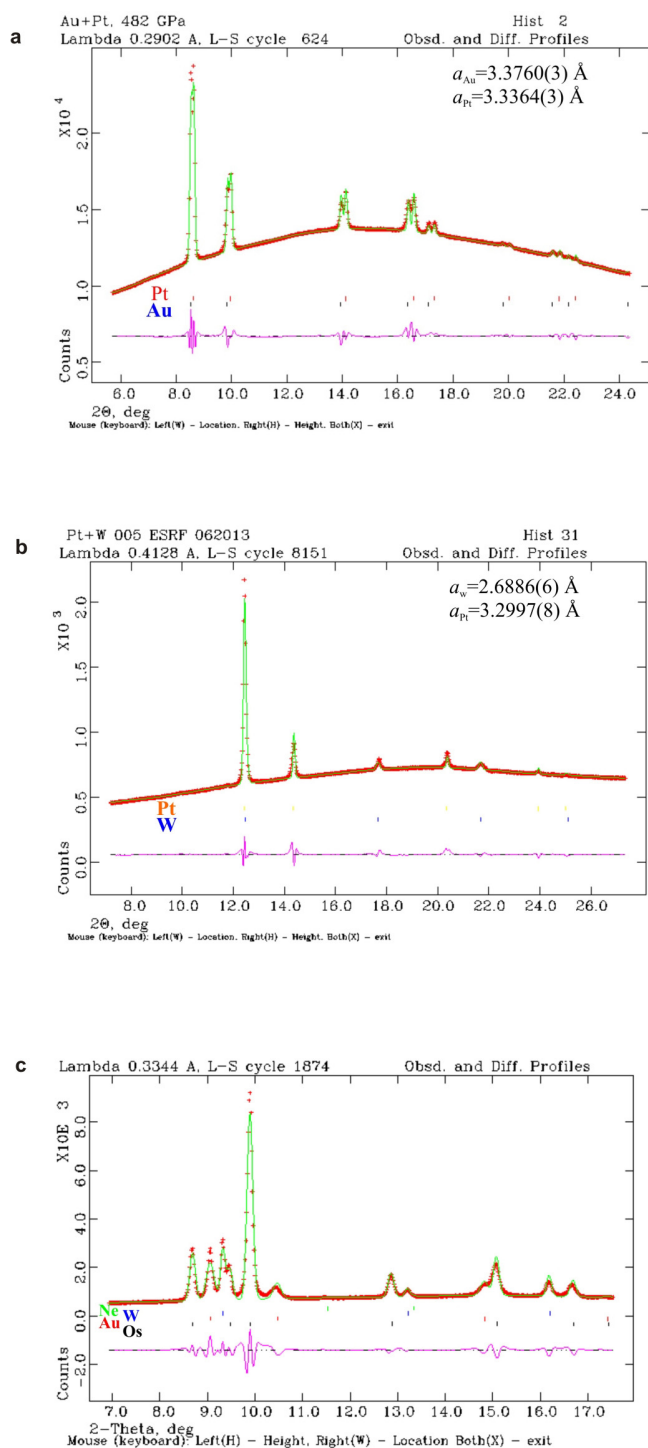
established in equation (4). Similar to the ETT, the anisotropy might be strengthened by a modification of the non-local pseudopotential, equations (1) and (2), acting on the valence electrons, owing to substantial reconstruction of inner  $5p$  and  $4f$  states at the transition. In the model given by equation (3), reconstruction of the inner states affects the electron-ion interaction via the ion core polarizability at imaginary frequencies. The derivation of equation (3) is based on the diagram consideration of the model of polarized ions in electron gas that is suggested in ref. 52. All the two-ion contributions to the total energy (that is, all diagrams containing two polarizability blocks) can be separated into two classes: one is the van der Waals interaction screened by itinerant electrons<sup>52</sup>; the other, as was demonstrated in ref. 50, can be interpreted as a second-order contribution of the pseudopotential, equation (3), with the only assumption being that the energy of excitation of the ionic core is higher than typical energies of the itinerant-electron subsystem (such as plasma frequency and the Fermi energy). More details can be found in refs 27 and 53.

31. Georges, A., Kotliar, G., Krauth, W. & Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
32. Boehnke, L. *et al.* Orthogonal polynomial representation of imaginary-time Green's functions. *Phys. Rev. B* **84**, 075145 (2011).
33. Aichhorn, M. *et al.* Dynamical mean-field theory within an augmented plane-wave framework: assessing electronic correlations in the iron pnictide LaFeAsO. *Phys. Rev. B* **80**, 085101 (2009).
34. Aichhorn, M., Pourovskii, L. & Georges, A. Importance of electronic correlations for structural and magnetic properties of the iron pnictide superconductor LaFeAsO. *Phys. Rev. B* **84**, 054529 (2011).
35. Gull, E. *et al.* Continuous-time Monte Carlo methods for quantum impurity models. *Rev. Mod. Phys.* **83**, 349–404 (2011).
36. Beach, K. S. D. Identifying the maximum entropy method as a special limit of stochastic analytical continuation. Preprint at <http://arxiv.org/abs/cond-mat/0403055> (2004).
37. Solov'ev, I. V., Dederichs, P. H. & Anisimov, V. I. Corrected atomic limit in the local-density approximation and the electronic structure of  $d$  impurities in Rb. *Phys. Rev. B* **50**, 16861–16871 (1994).
38. Czyżyk, M. T. & Sawatzky, G. A. Local-density functional and on-site correlations: the electronic structure of La<sub>2</sub>CuO<sub>4</sub> and LaCuO<sub>3</sub>. *Phys. Rev. B* **49**, 14211–14228 (1994).
39. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
40. Blaha, P., Schwarz, K., Madsen, G., Kvasnicka, D. & Luitz, J. *WIEN2k, An Augmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties* (Karlheinz Schwarz, Techn. Universität Wien, Austria, 2001).
41. Wills, J. M. *et al.* *Full-Potential Electronic Structure Method* (Springer, 2010).
42. Armiento, R. & Mattsson, A. E. Functional designed to include surface effects in self-consistent density functional theory. *Phys. Rev. B* **72**, 085108 (2005).
43. Mattsson, A. E. & Armiento, R. Implementing and testing the AM05 spin density functional. *Phys. Rev. B* **79**, 155101 (2009).
44. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996); erratum. **78**, 1396 (1997).
45. Haas, P., Tran, F. & Blaha, P. Calculation of the lattice constant of solids with semilocal functional. *Phys. Rev. B* **79**, 085104 (2009); erratum. **79**, 209902 (2009).
46. Ruban, A. V. & Abrikosov, I. A. Configurational thermodynamics of alloys from first-principles: effective cluster interactions. *Rep. Prog. Phys.* **71**, 046501 (2008).
47. Pourovskii, L. V. *et al.* Impact of electronic correlations on the equation of state and transport in  $\alpha$ -Fe. *Phys. Rev. B* **90**, 155120 (2014).
48. Katsnelson, M. I. & Trefilov, A. V. Fermi-liquid theory of electronic topological transitions and screening anomalies in metals. *Phys. Rev. B* **61**, 1643–1645 (2000).
49. Nizhankovskii, V. I. *et al.* Anisotropy of the thermal expansion of titanium due to proximity to an electronic topological transition. *JETP Lett.* **59**, 733–737 (1994).
50. Souvatzis, P., Eriksson, O. & Katsnelson, M. I. Anomalous thermal expansion in alpha-titanium. *Phys. Rev. Lett.* **99**, 015901 (2007).
51. Vaks, V. G. *et al.* Pre-transition softening and anomalous pressure dependence of shear constants in alkali and alkaline-earth metals due to band-structure effects. *J. Phys. Condens. Matter* **3**, 1409–1428 (1991).
52. Rehr, J. J., Zaremba, E. & Kohn, W. van der Waals forces in the noble metals. *Phys. Rev. B* **12**, 2062–2066 (1975).
53. Vonsovsky, S. V., Katsnelson, M. I. & Trefilov, A. V. Localized and itinerant behavior of electrons in metals. *Phys. Met. Metallogr.* **76**, 247–299 (1993).
54. Dewaele, A., Loubeyre, P. & Mezouar, M. Equations of state of six metals above 94 GPa. *Phys. Rev. B* **70**, 094112 (2004).
55. Litasov, K. D. *et al.* Thermal equation of state to 33.5 GPa and 1673 K and thermodynamic properties of tungsten. *J. Appl. Phys.* **113**, 133505 (2013).
56. Sahu, B. R. & Kleinman, L. Osmium is not harder than diamond. *Phys. Rev. B* **72**, 113106 (2005).
57. Dorfman, S. M., Prakapenka, V. B., Meng, Y. & Duffy, T. S. Intercomparison of pressure standards (Au, Pt, Mo, MgO, NaCl and Ne) to 2.5 Mbar. *J. Geophys. Res.* **117**, B08210 (2012); errata **117**, B11204 (2012).
58. Fei, Y. *et al.* Toward an internally consistent pressure scale. *Proc. Natl Acad. Sci. USA* **104**, 9182–9186 (2007).



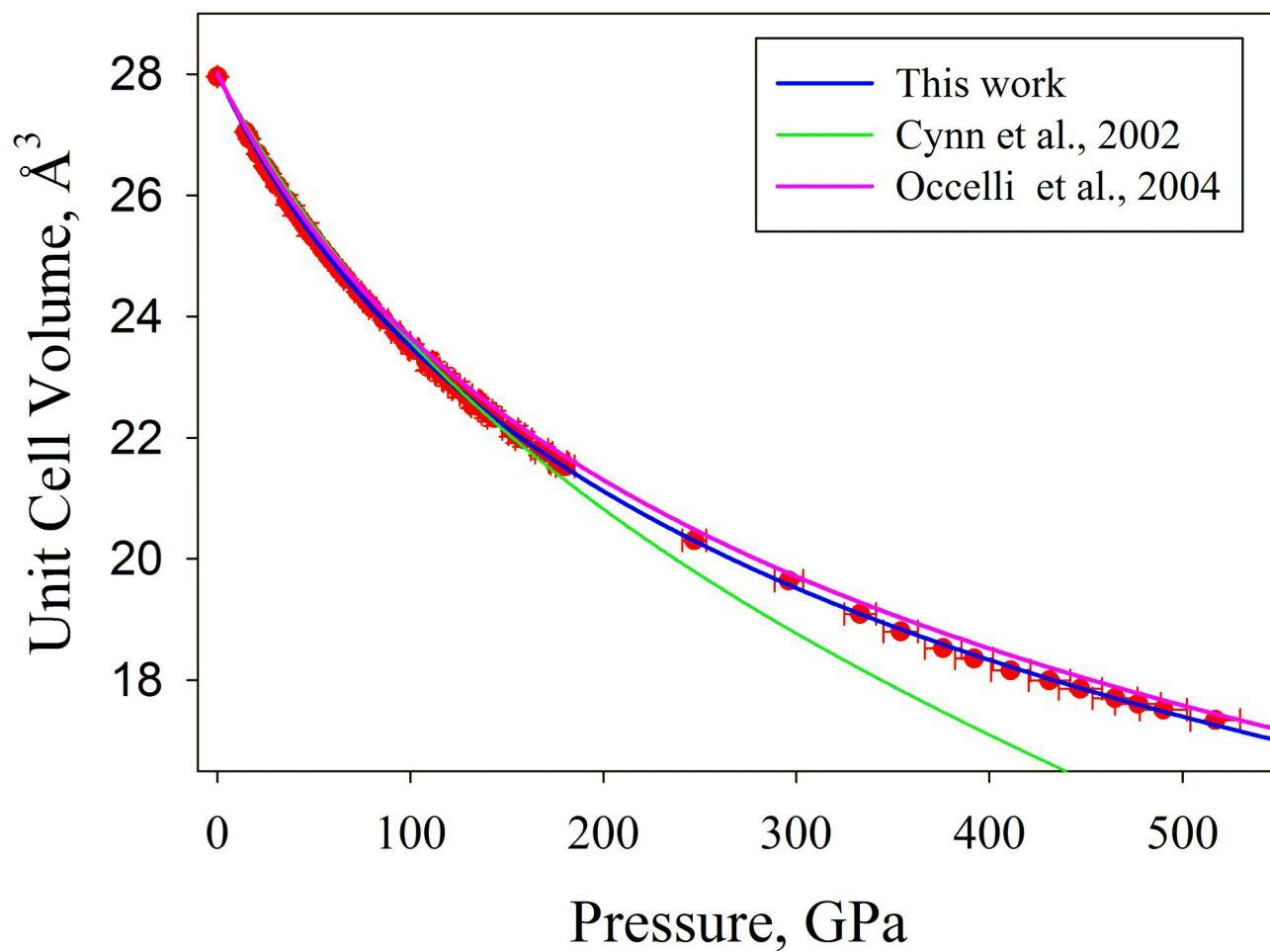
**Extended Data Figure 1 | Equations of state of W.** **a**, Comparison of different EOSs of W reported in Dewaele *et al.*<sup>54</sup>, Sokolova *et al.*<sup>23</sup>, and Litasov *et al.*<sup>55</sup>, and this work. Although the curves agree for pressures up to about 50 GPa, there is a substantial discrepancy for pressures around 0.5 TPa. **b**, Pressure dependence of the unit cell volume of W. Experimental data points (red solid

dots represent data collected using a Au–W mixture, green diamonds using a Pt–W mixture) were fitted using the third-order Birch–Murnaghan EOS (blue solid line,  $V_0 = 31.674(3) \text{ \AA}^3$  per unit cell,  $K_{300} = 307(2) \text{ GPa}$ ,  $K' = 4.53(4)$ ); data are equally well fitted with the Vinet EOS ( $V_0 = 31.686(7) \text{ \AA}^3$  per unit cell,  $K_{300} = 302(1) \text{ GPa}$ ,  $K' = 4.82(3)$ ).



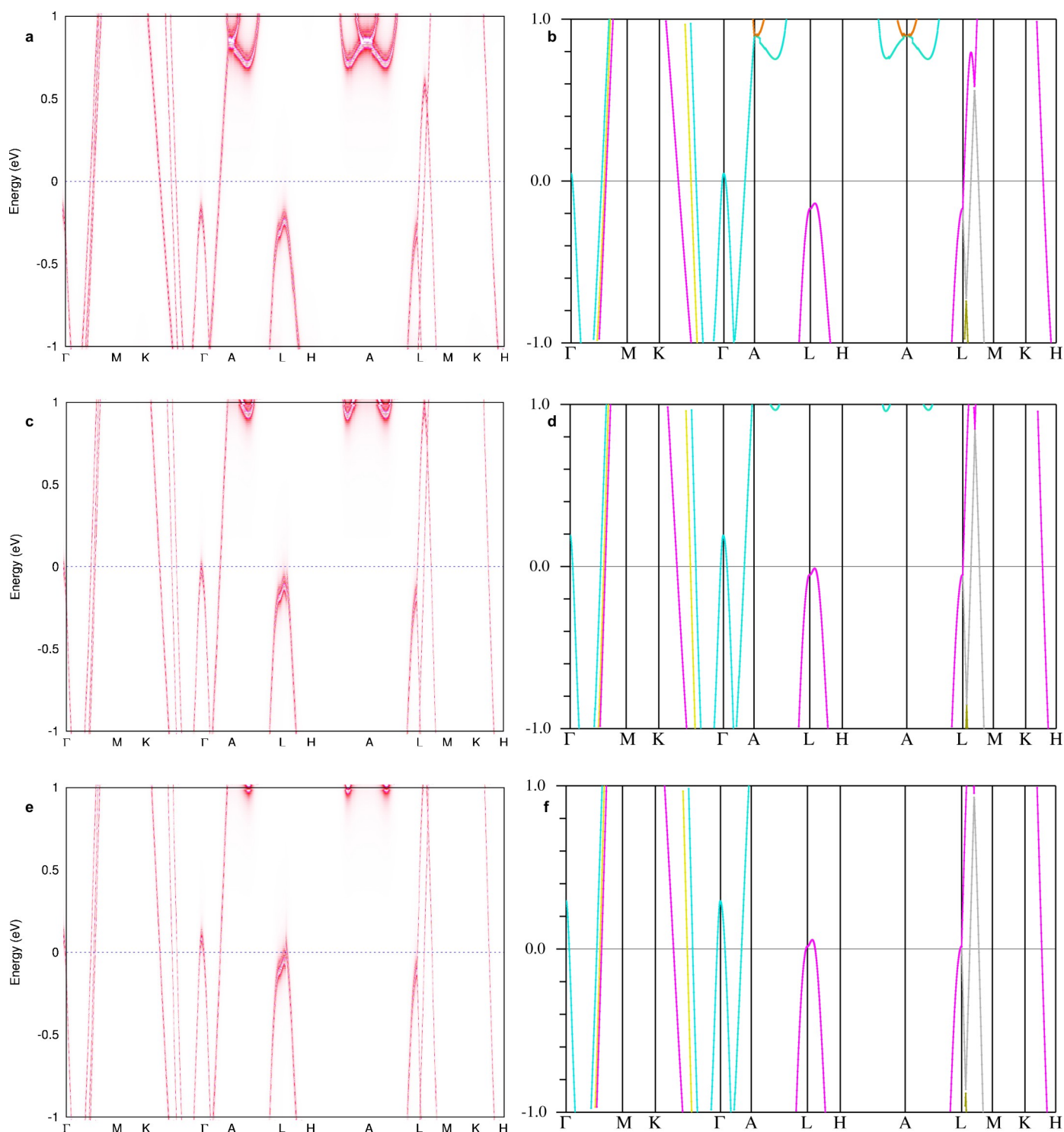
**Extended Data Figure 2 | Examples of the full-profile (GSAS) treated diffraction patterns.** **a**, Au–Pt mixture collected in a dsDAC at 482(5) GPa (Au pressure scale from ref. 20). Even at pressures in the proximity of 5 Mbar, powder X-ray diffraction data are sufficient to clearly resolve peaks of Au and Pt, and accurately determine lattice parameters of both metals. **b**, W–Pt mixture, unsuccessful experiment in a dsDAC. According to the W EOS, the pressure is 461(7) GPa, whereas according to the Pt EOS, it is 559(10) GPa. This inconsistency in pressures is a result of an inhomogeneous distribution of the two metals between secondary anvils in the dsDAC; such data cannot be used for constraining EOSs. This observation also shows that very large pressure gradients are possible in dsDACs. **c**, Mixture of Os ( $a = 2.5404(6) \text{ \AA}$ ,  $c = 4.0386(10) \text{ \AA}$ ), W ( $a = 2.9133(5) \text{ \AA}$ ), and Au ( $a = 3.6657(6) \text{ \AA}$ ) collected in a conventional DAC (Ne used as the pressure transmitting medium) at a pressure of 134(2) GPa.





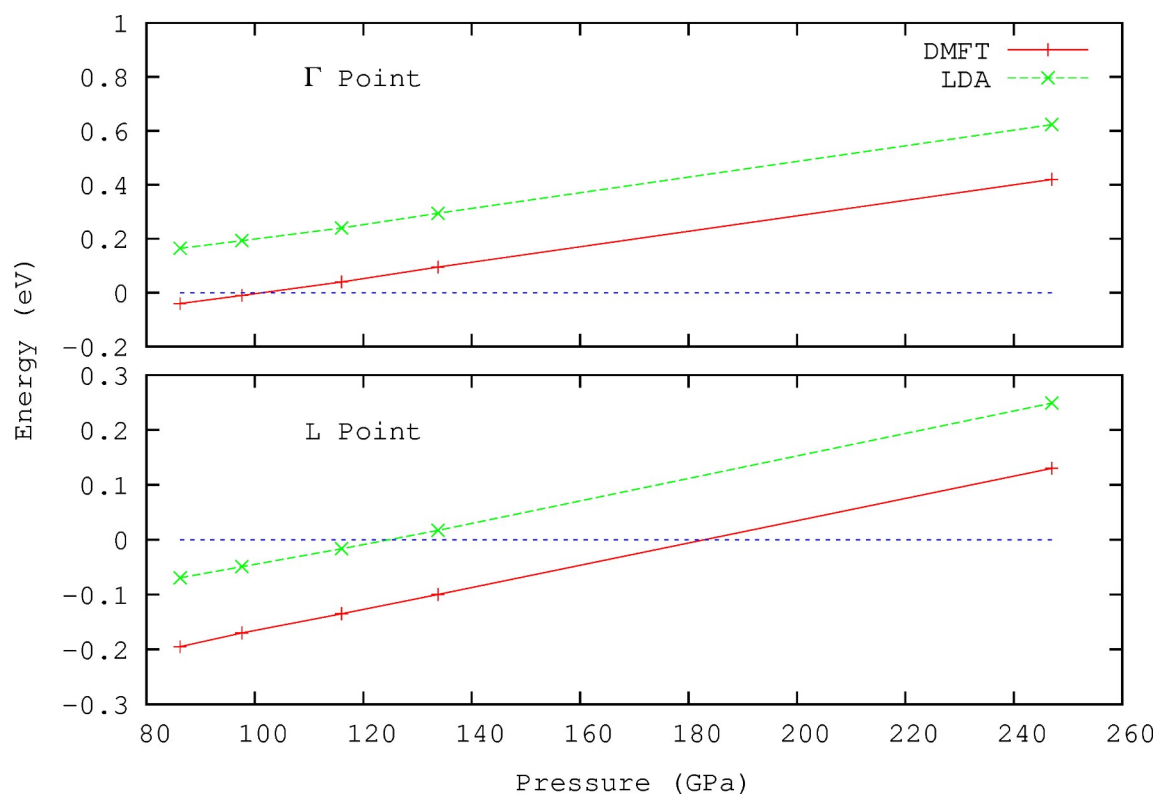
**Extended Data Figure 3 | Experimental dependence of the unit cell volume of Os on pressure in comparison with EOSs reported in the literature.** The magenta line is from Occelli *et al.*<sup>3</sup> ( $K_{300} = 421$  GPa,  $K' = 4.0$ ) and the green line is from Cynn *et al.*<sup>4</sup> ( $K_{300} = 463$  GPa,  $K' = 2.8$ ). These

data have been re-fitted using the ruby pressure scale as suggested in refs 20, 22 and references therein. Experimental data points (solid red dots) were fitted using the third-order Birch–Murnaghan EOS (blue line,  $V_0 = 28.02(4)$   $\text{\AA}^3$  per unit cell,  $K_{300} = 399(6)$  GPa,  $K' = 4.04(4)$ ).



**Extended Data Figure 4 | Electronic band structure of Os at moderate compressions along the high-symmetry lines in the hcp Brillouin zone.** Energies are given relative to the Fermi energy, which is taken to be zero. Calculations are carried out at pressures of 0 GPa (a, b), 97.5 GPa (c, d), and 134 GPa (e, f). a, c, e, The  $k$ -resolved spectral functions  $A(\mathbf{k}, \omega)$  obtained with LDA+DMFT. b, d, f, The band structure obtained with LDA. In both cases, we used the experimental lattice parameters ( $a = 2.734 \text{ \AA}$ ,  $c/a = 1.580$  at 0 GPa;  $a = 2.578 \text{ \AA}$ ,  $c/a = 1.589$  at 97.5 GPa;  $a = 2.540 \text{ \AA}$ ,  $c/a = 1.590$  at

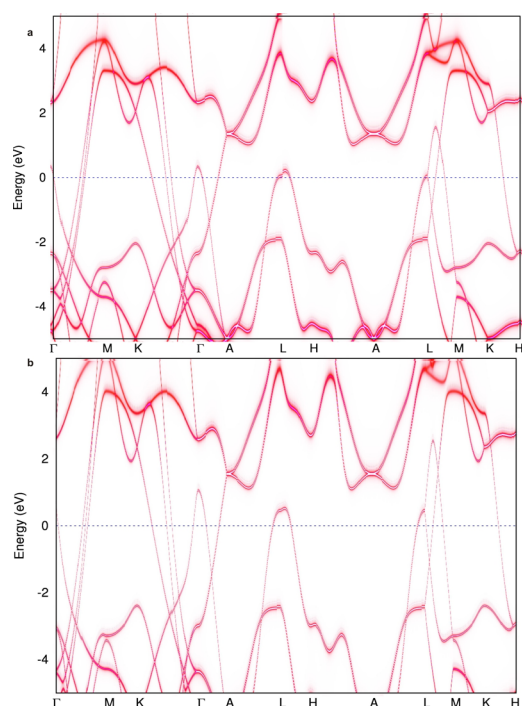
134 GPa). Our LDA+DMFT calculations predict that two ETTs occur in hcp Os upon compression. In a, a band at the  $\Gamma$  point is well below the Fermi energy at ambient pressure; however, it nearly touches the Fermi energy at 97.5 GPa, and the corresponding hole pocket has already appeared at 134 GPa, giving rise to a change of the Fermi surface topology, that is, to an electronic topological transition at about 101.5 GPa (Extended Data Fig. 5). Our LDA+DMFT calculations also indicate that another ETT at the L point should occur above 134 GPa, at about 183 GPa (Extended Data Fig. 5).



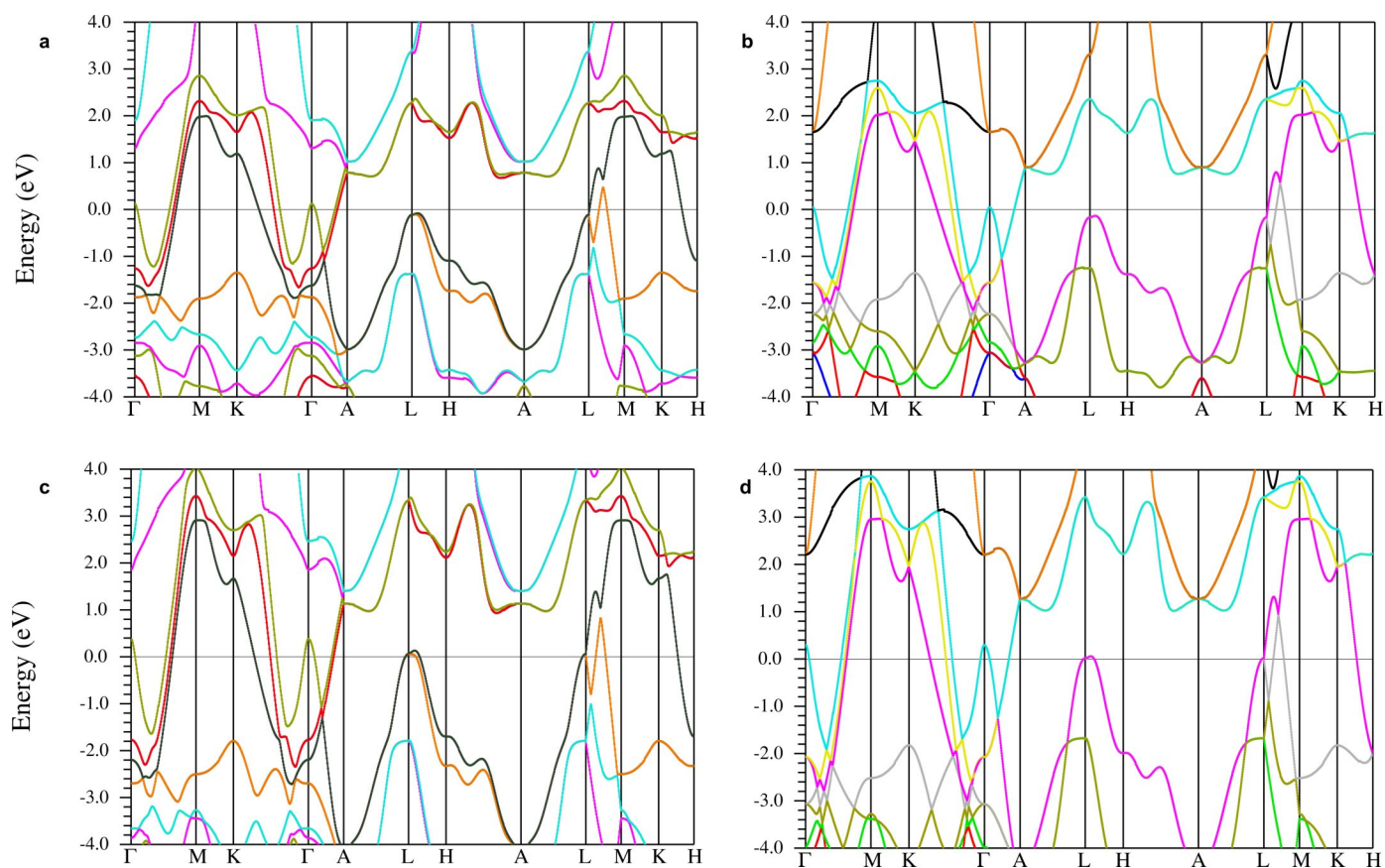
**Extended Data Figure 5 | The position of the relevant bands at the  $\Gamma$  (top) and L (bottom) points obtained from LDA+DMFT and LDA calculations as a function of pressure.** The positive values indicate appearance of the corresponding hole pockets. Energies are given relative to the Fermi energy. A closer examination of the band shape in the vicinity of the L point reveals that this hole pocket first appears along the L–H line and then extends to include

the L point. To estimate more precisely the critical pressures for the ETTs, we plot the positions of the relevant bands at the  $\Gamma$  and L points with respect to the Fermi level and obtain the values of the critical pressures by interpolation. The LDA+DMFT calculations predict that the hole pockets at the  $\Gamma$  and L points appear at about 101.5 GPa and 183 GPa, respectively.



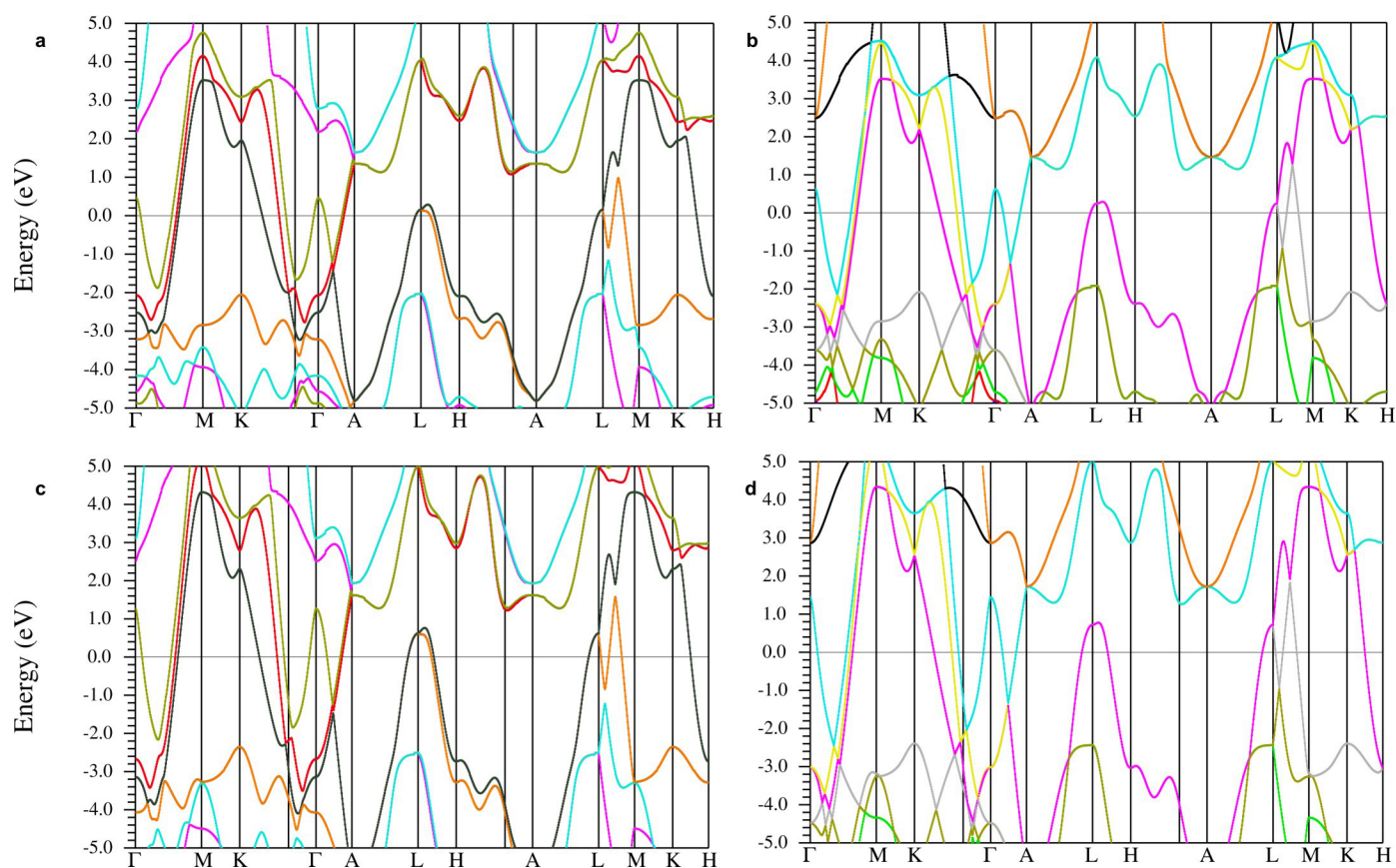


**Extended Data Figure 6 |  $k$ -resolved spectral functions  $A(k, \omega)$  of Os at ultra-high compressions along the high-symmetry lines in the hcp Brillouin zone.** **a, b,** Calculations are carried out at pressures of 247 GPa (**a**) and 477 GPa (**b**) using LDA+DMFT. The experimental lattice parameters were used in the calculations ( $a = 2.449$  Å,  $c/a = 1.596$  at 247 GPa;  $a = 2.335$  Å,  $c/a = 1.597$  at 477 GPa). Energies are given relative to the Fermi energy, which is taken to be zero. Hole pockets are present at the  $\Gamma$  and L points at  $P = 247$  GPa. Increasing the pressure to 477 GPa does not induce any new ETTs. Moreover, no features of the band structure suggest that new ETTs might be induced by further increase of the pressure within reasonable limits.



**Extended Data Figure 7 | Electronic band structure of Os at moderate compressions along the high-symmetry lines in the hcp Brillouin zone.** a–d, Band structure of Os at 0 GPa (a, b) and at 134 GPa (c, d). Energies are given relative to the Fermi energy. In a and c, spin–orbit coupling has been

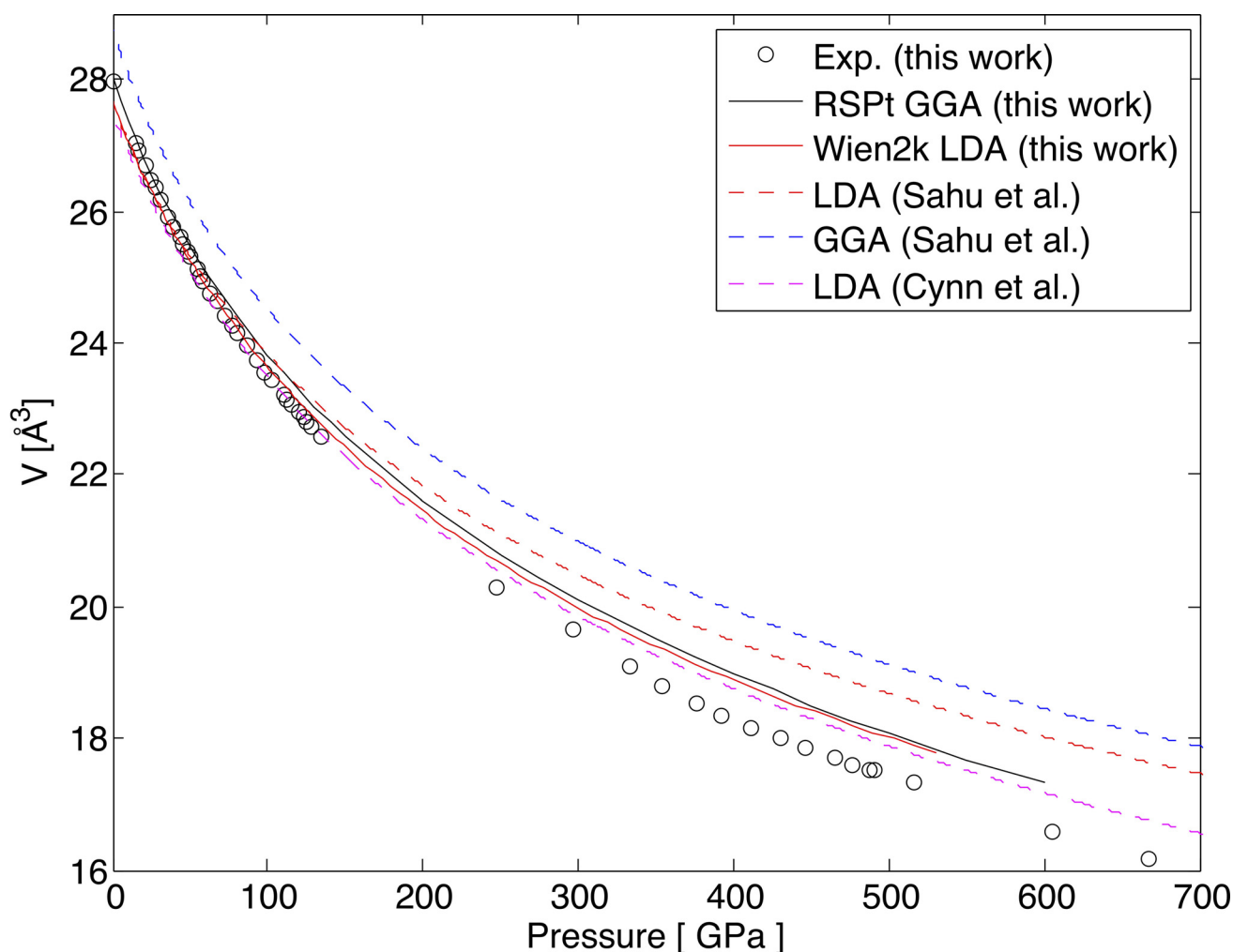
included, whereas in b and d, the energies were obtained with the scalar-relativistic approximation. Experimental lattice parameters were used in the calculations (see Extended Data Fig. 4). Bands are shown in different colours for clarity.



**Extended Data Figure 8 | Electronic band structure of Os at extreme compressions along the high-symmetry lines in the hcp Brillouin zone.** **a–d**, Band structure of Os at 247 GPa (**a**, **b**) and at 477 GPa (**c**, **d**). Energies are given relative to the Fermi energy. In **a** and **c**, spin-orbit coupling has been

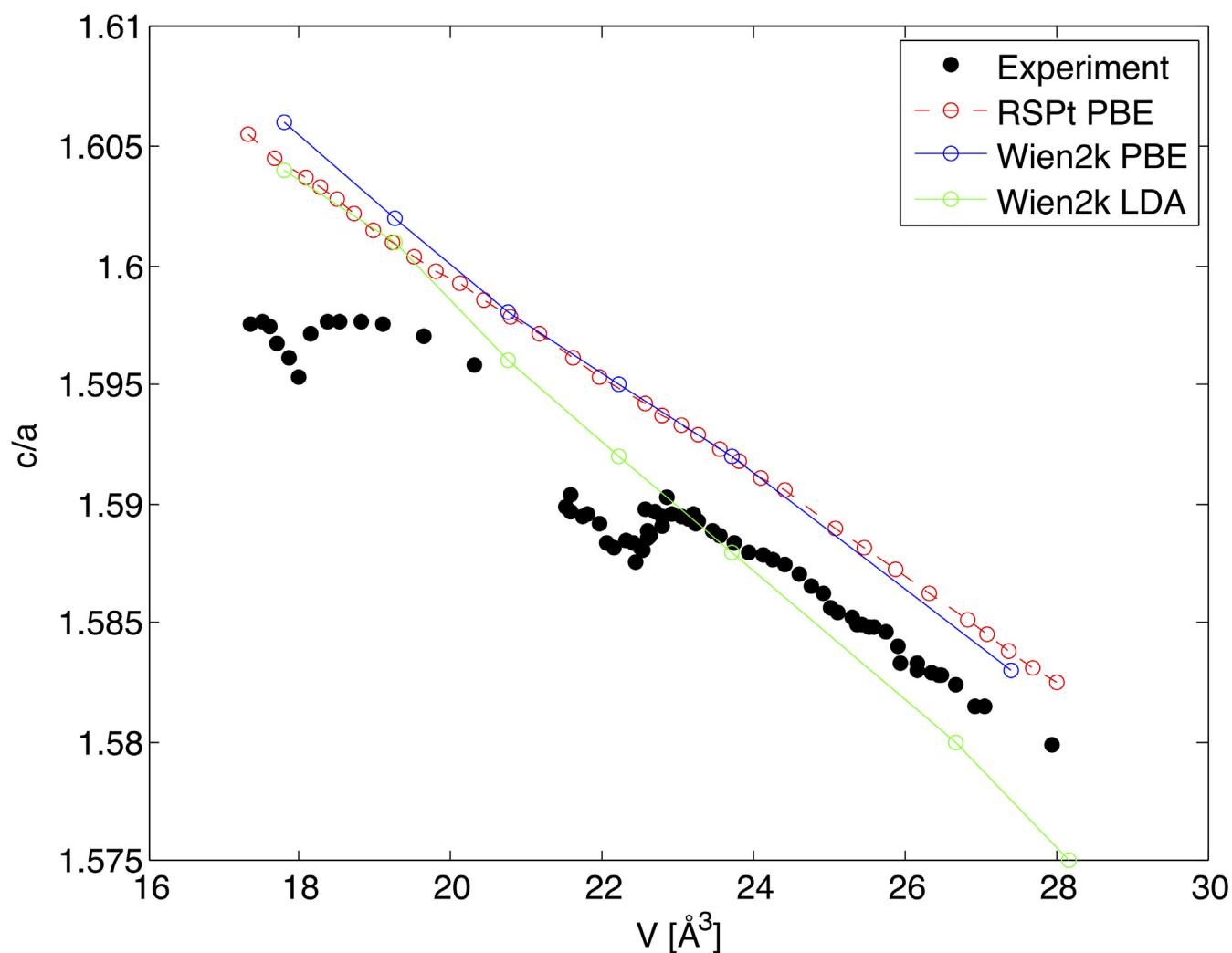
included, whereas in **b** and **d**, the energies were obtained with the scalar-relativistic approximation. Experimental lattice parameters were used in the calculations (see Extended Data Fig. 6).





**Extended Data Figure 9 | Comparison of the EOS for hcp Os calculated using different approximations within DFT with the experimental EOS measured in this work.** Shown are GGA calculations carried out by us using the RSPt method (black solid line) and by Sahu *et al.*<sup>56</sup> (blue dashed line), as well as LDA calculations carried out by us using the Wien2k method (red solid line), by Cynn *et al.*<sup>4</sup> (pink dashed line), and by Sahu *et al.*<sup>56</sup> (red dashed line). The experimental EOS obtained in this work is shown with black circles. The different curves agree reasonably well for pressures up to about 50 GPa. In contrast, there is noticeable discrepancy at ultra-high pressure, and there is no theoretical EOS that accurately describes the experiment for pressures around 0.5 TPa. Local and semi-local approximations within DFT are insufficient to describe the  $P$ - $V$  relationship of Os under extreme conditions. Part of the reason for the disagreement between theory and experiment might be related to the improper account of the many-electron effects within the theory. In principle, LDA and GGA work better at high pressure; however, errors that might be important at low pressure<sup>47</sup> could propagate through the EOS to the whole pressure range, owing to the use of all the calculated points in the fitting of the energy versus volume data by, for example, the third-order Birch–Murnaghan EOS used in this work. To justify this statement, consider the calculated EOS parameters, summarized in Extended Data Table 1. The

equilibrium volumes calculated by us differ from the experiment (Extended Data Table 1) by less than 1.5%. On the other hand, the overestimation of the calculated bulk moduli ( $B$ ) and their pressure derivatives ( $B'$ ) is greater, about 10%. Our EOS parameters are within the range of theoretical parameters available in the literature, which are fitted for  $P < 100$  GPa (Extended Data Table 1). We deal with a highly incompressible metal, for which typical DFT errors in  $B$ , and especially in  $B'$ , translate into large differences in  $P$ - $V$  relationships at ultra-high pressure. Even in this regime, the error in volume at a fixed pressure remains within typical DFT limits of about 2%–3%. However, the pressure calculated at fixed volume can differ by several tens of gigapascals. This difference is due to errors in  $B$  and  $B'$  calculated at ambient pressure, coupled to a very high value of  $B$ . The use of more advanced theoretical methods could improve the calculated EOS. In ref. 47, a substantial reduction of  $B$  in isoelectronic hcp Fe is demonstrated using a LDA+DMFT approach. Here the effect is expected to be smaller, but may still be sufficient to improve the agreement with experiment. Our results demonstrate a need to further develop the electronic structure theory, with the experiment reported here providing a bench-mark for the theory. On the other hand, we consistently used experimental lattice parameters in the discussion of the electronic structure of Os in this study.



**Extended Data Figure 10 | Comparison of the ratio of the lattice parameters  $c/a$  for hcp Os.** The ratio is calculated using the RSPt method at  $T = 0$  K with a PBE-GGA (red dashed line), and using Wien2k with a LDA (green line) and a PBE-GGA (blue line). The room-temperature experimental results obtained in this work are shown with filled black dots. Agreement between the calculated and experimental  $c/a$  ratio is typical for DFT calculations. The theoretical results do not show any peculiarity of the lattice-parameter ratio;

however, the calculations are carried out at  $T = 0$  K, whereas the experimental data are taken at room temperature. Therefore, a direct comparison between theory and experiment is nontrivial (Methods). In hcp metals, the effect of the electronic transitions on the lattice-parameter ratio should become visible at finite temperatures, owing to the peculiarities of the thermal expansion coefficients, which are anisotropic along different directions of the crystal lattice.

Extended Data Table 1 | EOS parameters of Au, Pt, W, and Os

a	Metal	3 <sup>rd</sup> order Birch–Murnaghan EOS			Vinet EOS			Reference; Range
		$V_0$ , Å <sup>3</sup> /unit cell	$K_{300}$ , GPa	$K'$	$V_0$ , Å <sup>3</sup> /unit cell	$K_{300}$ , GPa	$K'$	
Au		67.85	167.5	5.61				20:550 GPa
					67.85	167	5.88	57:250 GPa
Pt		60.389(5)	274(2)	5.23(3)	60.381(5)	261(2)	5.83(3)	58:100 GPa
								<b>This work</b>
Pt		60.38	276.4	5.12				480 GPa
					60.38	277	5.08	20:550 GPa
W		31.674(3)	306.8(1.5)	4.53(4)	31.686(7)	301.9(1.2)	4.82(3)	54:94 GPa
								57:250 GPa
W					31.724	296	4.30	200 GPa
					31.691	317	3.16	54:94 GPa
Os		28.02(4)	398.5(5.9)	4.04(4)	28.08(5)	380(7)	4.48(6)	55:33 GPa
					31.72	308	4.25	23:400 GPa
Os		27.96	462	2.4				<b>This work</b>
		27.941	411	4.0				517 GPa
Os		27.977	395	4.5				4:65 GPa
								3:75 GPa
								2:58 GPa

b				
		V [Å <sup>3</sup> ]	B [GPa]	B'
GGA (RSPt, this work)		28.00	448	4.29
		28.00	423	4.67
GGA (RSPt, this work, fitted at P<100 GPa)		28.72	401	5.47
			382	4.60
GGA (Ref. 3)		27.66	447	4.41
			437	4.46
LDA (Wien2k, this work)		27.60	454	5.26
		27.50	445	4.4
LDA (Ref. 3)				
LDA (Ref. 56)				
LDA (Ref. 4)				

**a**, Experimental data; **b**, theoretical data. Values in **a** are from this work and refs 2–4, 20, 23, 54, 55, 57, and 58. Presented in **b** are GGA calculations carried out by us using the RSPt method and the calculations presented in ref. 56, as well as LDA calculations carried out by us using the Wien2k method and the calculations presented in refs 3, 4, and 56.



# Computational design of co-assembling protein–DNA nanowires

Yun Mou<sup>1</sup>, Jiun-Yann Yu<sup>2</sup>, Timothy M. Wannier<sup>2</sup>, Chin-Lin Guo<sup>3</sup> & Stephen L. Mayo<sup>1,2</sup>

Biomolecular self-assemblies are of great interest to nanotechnologists because of their functional versatility and their biocompatibility<sup>1</sup>. Over the past decade, sophisticated single-component nanostructures composed exclusively of nucleic acids<sup>2–5</sup>, peptides<sup>6–8</sup> and proteins<sup>9–15</sup> have been reported, and these nanostructures have been used in a wide range of applications, from drug delivery<sup>16</sup> to molecular computing<sup>17</sup>. Despite these successes, the development of hybrid co-assemblies of nucleic acids and proteins has remained elusive. Here we use computational protein design to create a protein–DNA co-assembling nanomaterial whose assembly is driven via non-covalent interactions. To achieve this, a homodimerization interface is engineered onto the *Drosophila* Engrailed homeodomain (ENH), allowing the dimerized protein complex to bind to two double-stranded DNA (dsDNA) molecules. By varying the arrangement of protein-binding sites on the dsDNA, an irregular bulk nanoparticle or a nanowire with single-molecule width can be spontaneously formed by mixing the protein and dsDNA building blocks. We characterize the protein–DNA nanowire using fluorescence microscopy, atomic force microscopy and X-ray crystallography, confirming that the nanowire is formed via the proposed mechanism. This work lays the foundation for the development of new classes of protein–DNA hybrid materials. Further applications can be explored by incorporating DNA origami, DNA aptamers and/or peptide epitopes into the protein–DNA framework presented here.

The functionality of single-component biomolecular materials is ultimately limited by the physical and/or chemical nature of their constituent parts. Advanced materials that integrate a variety of chemical building blocks will facilitate the incorporation of greater functionality in materials design<sup>18,19</sup>. Hybrid assemblies that integrate both protein and DNA have been engineered for several useful applications, including catalytic cascades<sup>20</sup>, amplified biosensors<sup>21</sup> and functional templates for the growth of inorganic materials<sup>22</sup>. These materials, however, rely on DNA self-assembly to form a scaffold to which proteins are attached via chemical conjugation methods<sup>20–22</sup>—an approach that is problematic for two reasons. First, chemical conjugation as a means to molecular assembly has several drawbacks, including a complicated synthesis process, heterogeneous labelling problems and inapplicability to *in vivo* systems. Conjugation problems have been mitigated somewhat through the use of RNA aptamers, which allow the programmed positioning of proteins *in vivo*<sup>23</sup>. Nevertheless, existing protein–nucleic-acid materials still suffer from a second shortcoming in that the main body of the nanostructure is formed solely from nucleic acids. A co-assembling scaffold that integrally incorporates both DNA and protein would allow greater structural diversity and permit easy temporal and locational control over the self-assembly. However, the construction of such a hybrid has thus far remained elusive.

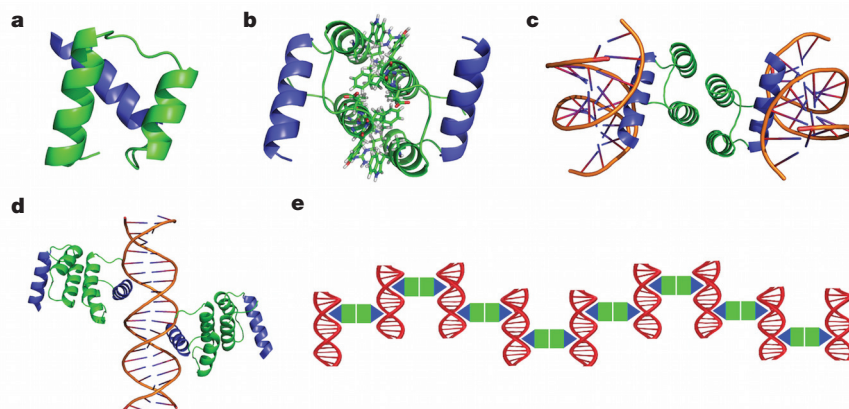
Here we create co-assembling protein–DNA nanomaterials via non-covalent interactions. We start by engineering a dual-function protein that contains both a homodimerization domain and a DNA-binding domain. This designed protein homodimer serves as a

primary building block that binds two DNA molecules; each monomeric subunit binds a specific fragment of dsDNA. We chose ENH as our scaffold protein for the following reasons: (1) a single-helical peptide from ENH (helix 3) binds a target dsDNA motif (TAATNN)<sup>24</sup> as tightly as the full-length protein (dissociation constant ( $K_d$ ) is in the nanomolar range)<sup>25</sup>, indicating that DNA-binding functionality could be isolated to a specific domain; (2) ENH has been intensely studied using computational tools, and highly stable variants<sup>26</sup>, full-sequence redesigns<sup>27</sup> and a *de novo* homodimer design<sup>28</sup> have been generated; and (3) as a three-helix protein, ENH provides a surface for homodimer interface design (the exterior faces of helices 1 and 2) that is structurally opposite its DNA-binding helix 3.

Figure 1 illustrates our protein–DNA nanomaterial design strategy. Using the *Drosophila melanogaster* ENH crystal structure (Protein Data Bank (PDB) accession 1ENH) as our docking subunit (Fig. 1a), we performed fast Fourier transform-based docking to generate C2 symmetrical homodimer models. The best model exhibited parallel intermolecular helical packing between helices 1 and 2 of each of the ENH monomers. Computational protein design was then used to design the interface residues of the docked model to minimize the free energy of the intermolecular side-chain interactions (Fig. 1b). Early design variants were characterized and iteratively improved with the use of a molecular dynamics screening protocol<sup>29</sup>. We named the final designed variant dualENH because it has dual functionality: it can both homodimerize and bind dsDNA. A dualENH homodimer serves as the protein building block for nanomaterial assembly because it has two binding sites for dsDNA (one on each of two opposite faces of the homodimer), as shown in an aligned model (Fig. 1c). The second designed component of the nanomaterials is a dsDNA building block with protein-binding sites variously placed along the double helix (Fig. 1d and Extended Data Fig. 1a). By tuning the positioning of binding sites on the dsDNA and then simply mixing the two designed components (DNA and protein) together, we were able to achieve co-assembly of both irregularly shaped particles of protein and DNA (Extended Data Fig. 1b) and well-ordered protein–DNA nanowires (Fig. 1e).

We characterized dualENH to confirm that: (1) it forms a homodimer with helical secondary structure and no high-order aggregation; (2) it binds dsDNA probes specifically; and (3) each homodimer binds two dsDNA molecules (Supplementary Information and Extended Data Figs 2–4). We next sought to observe the protein–DNA self-assembly using fluorescence microscopy. Figure 2a shows that nanoparticles were formed immediately after 5  $\mu$ M dualENH was mixed with 2  $\mu$ M (TAA)<sub>5</sub>. The particles were irregularly shaped with diameters of up to several micrometres (see Extended Data Fig. 5a for the size distribution). The irregularity of shape was expected because (TAA)<sub>5</sub> has four ENH-binding sites (TAATAA) that each face in a different direction off of the dsDNA helix, which causes particle growth to occur in a random branching pattern (Extended Data Fig. 1b). The particles are invisible under bright-field microscopy (Extended Data Fig. 5b) because of the transparency of protein and

<sup>1</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. <sup>2</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA. <sup>3</sup>Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125, USA.



**Figure 1 | Protein-DNA nanomaterial design strategy.** **a**, Helix 1 and helix 2 (green) of ENH were engineered into a homodimerization domain, and helix 3 (blue) is the native DNA-binding domain. **b**, The interface of the docked model was designed for homodimerization. **c**, The designed homodimer, named dualENH, binds two dsDNA fragments on its outward faces. This model was generated by aligning the homodimer model in **b** with the

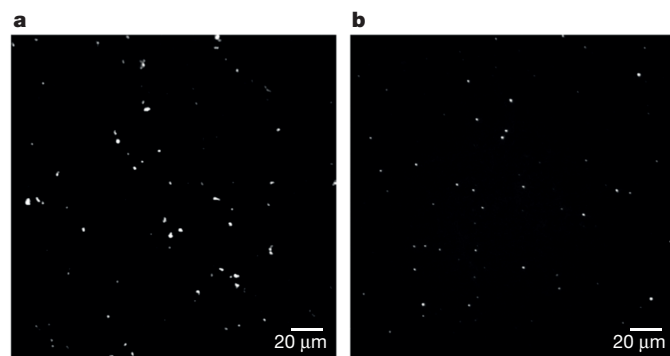
ENH-DNA co-crystal structure (PDB accession 3HDD). **d**, Two protein-binding sites were engineered onto a dsDNA fragment so that two dualENH dimers would bind  $180^\circ$  apart along the double helix. **e**, The dualENH protein in **c** and the dsDNA fragment in **d** co-assemble into a protein-DNA nanowire. Note that this two-dimensional cartoon is for purposes of illustration only, and that the three-dimensional design model of the nanowire is spiralled.

DNA to visible light. A control experiment showed that a solution of  $(\text{TAA})_5$  by itself did not form any particles (Extended Data Fig. 5c). When lower concentrations of dualENH (500 nM) and  $(\text{TAA})_5$  (200 nM) were used, a smaller and more uniform particle distribution was observed (Extended Data Fig. 5d). Further decreasing the protein concentration ( $<200$  nM) greatly reduced the number of particles formed (Extended Data Fig. 5e). This reduction may be due to dissociation of the homodimer at low concentrations. To confirm that nanoparticle formation occurs via the proposed mechanism (Fig. 1e), we designed a particle inhibition experiment using dsDNA with a single binding site as the inhibitor. The single binding site on these dsDNAs should terminate particle growth. Pre-incubation of 500 nM dualENH with only trace amounts of single-binding-site dsDNA (5 nM) abolished particle formation completely when 200 nM  $(\text{TAA})_5$  was added (Extended Data Fig. 5f).

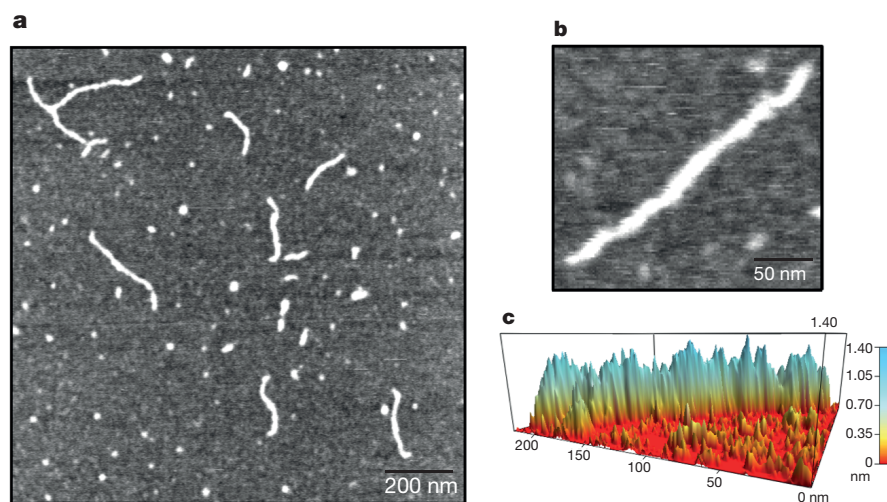
To form a linear protein-DNA co-assembly, as illustrated in Fig. 1e, the dsDNA building block must have two protein-binding sites about  $180^\circ$  apart on the dsDNA double helix (Fig. 1d). We designed a 25-nucleotide dsDNA molecule with an 11-nucleotide binding motif

(TAATTTAATTT, named motif 11) that contains two ENH-binding motifs (TAATTT) facing in opposite directions off of the helix. At the same protein and DNA concentrations used in the earlier particle-forming experiment (5  $\mu\text{M}$  and 2  $\mu\text{M}$ , respectively) (Fig. 2a), dsDNA containing motif 11 and dualENH formed much smaller and more uniform particles, with none growing greater in size than the diffraction limit (submicrometre) (Fig. 2b). The reduced particle size may be a result of fewer protein-binding sites on the DNA building block (two versus four). Fewer protein-binding sites would decrease the entropy of self-assembly and increase the chance of binding-site poisoning. We used atomic force microscopy (AFM) to study the topology of the protein-DNA co-assembly formed with dualENH and the two-binding-site dsDNA. Nanowire structures were clearly observed with a width of  $\sim 15$  nm and a length of up to  $\sim 300$  nm (Fig. 3a, b), which corresponds to  $\sim 60$  repeated units of protein and dsDNA on the basis of the design model. In accordance with the design model, the observed width of the nanowires ( $\sim 15$  nm) is consistent with the length of the dsDNA ( $\sim 9$  nm), considering that AFM usually overestimates the length in the  $x$ - $y$  plane due to the size of the tip. The height of the nanowire is  $\sim 1.0$  nm (Fig. 3c), which is on the order of the diameter of a dsDNA fragment ( $\sim 2$  nm); the decreased height could be due to compression by the hard AFM tip ( $k = 3 \text{ N m}^{-1}$ ). Note that dsDNA molecules lie flat on a  $\text{Mg}^{2+}$ -mica substrate due to strong electrostatic interactions between dsDNA and  $\text{Mg}^{2+}$  ions, and thus a flat ribbon, instead of the spiralled structure predicted in the design model, was observed in the AFM measurements.

We solved the co-crystal structure of dualENH with a dsDNA probe containing motif 11 (Fig. 4). These structural data confirmed both the spiralled nature of the nanowire and the dual functionality of dualENH: (1) dualENH uses helix 3 to bind dsDNA just as wild-type ENH does; and (2) dualENH uses the surfaces of helix 1 and helix 2 to form a homodimer (Fig. 4a). However, the co-crystal structure reveals two homodimer configurations of dualENH that differ slightly from each other, as seen by their backbone root mean squared deviation (r.m.s.d.) of  $4.0 \text{ \AA}$  (Fig. 4b). This unexpected result might be caused by crystal packing forces, especially as each dsDNA molecule in the crystal forms a superhelix (see end-to-end packing of dsDNA fragments in Extended Data Fig. 6a). Given this observation, we cannot conclude that either of the observed dualENH dimers in the crystal structure reflects the predominant dimer structure in solution. The two dualENH dimer crystal structures have backbone r.m.s.d. values to the design model of  $3.8 \text{ \AA}$  and  $3.9 \text{ \AA}$ , respectively (see Extended Data Fig. 6b, c). The co-crystal structure confirms that dualENH binds the dsDNA with its designed 11-nucleotide binding motif; however, it



**Figure 2 | Fluorescence microscopy of protein-DNA nano-objects.** **a**, dsDNA  $(\text{TAA})_5$  fragments were labelled with the fluorescent dye Cy3. A fluorescent image was taken of particles formed by mixing 5  $\mu\text{M}$  dualENH with 2  $\mu\text{M}$  Cy3- $(\text{TAA})_5$  in 20 mM Tris-HCl and 100 mM NaCl at pH 8.0. The particles formed irregular shapes up to  $\sim 5 \mu\text{m}$  in diameter. **b**, Same experiment as in **a** except that 25-nucleotide dsDNA fragments containing motif 11 (TAATTTAATTT) in the middle (CGAGTGTAATTTAATTTTCCTCGAC; highlighted in bold) were used instead of  $(\text{TAA})_5$  fragments. All particle sizes are under the diffraction limit (submicrometre). The shapes of the particles are slightly oval instead of being symmetrical (circular) due to moderate geometrical aberrations of the microscopy system.



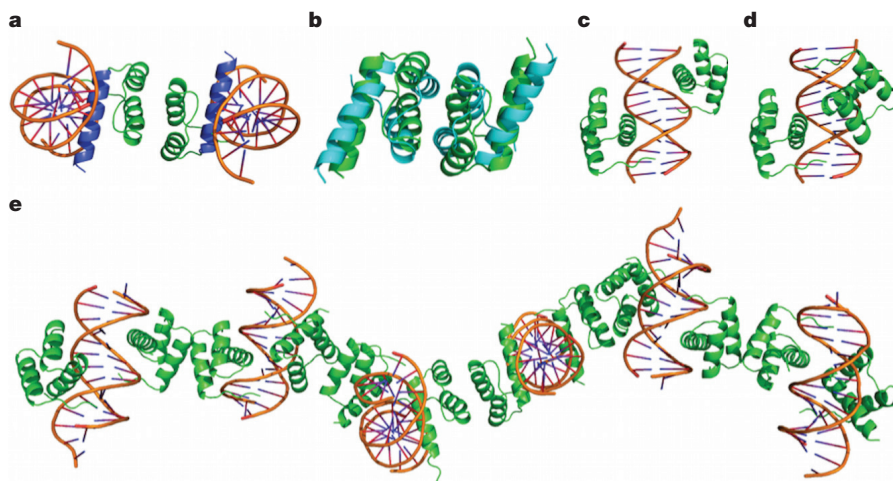
**Figure 3 | Atomic force microscopy of protein–DNA nanowires.** **a**, Representative AFM image obtained after mixing 5  $\mu$ M dualENH with 2  $\mu$ M of the two-binding-site dsDNA (25-nucleotide dsDNA containing motif 11, as in Fig. 2b). Nanowire structures  $\sim$ 15 nm wide and up to 300 nm long are clearly

visible. **b**, Magnified image of a single nanowire  $\sim$ 250 nm in length. **c**, Three-dimensional topology display of **b** shows that the height of the nanowire is  $\sim$ 1.0 nm.

also reveals two configurations of protein–DNA binding. One configuration is consistent with our design model, in which two dualENH molecules bind to the 11-nucleotide motif exactly opposite each other on the dsDNA double helix (Fig. 4c). The other configuration, however, features one of the dualENH bindings in an inverted orientation; that is, it binds to the reverse complementary sequence (AAATTA) of the optimal binding motif (TAATTT) (Fig. 4d). This suboptimal binding has also been seen in other ENH crystal structures, presumably due to the high concentrations of protein and DNA used for crystallization. Because of this alternative protein–DNA-binding configuration, the nanowire in the co-crystal structure is slightly kinked; nevertheless, an infinitely repeated protein–DNA nanowire is observed (Fig. 4e). Future work should be focused on gaining greater control over geometric specificity, especially in efforts to engineer more sophisticated three-dimensional structures.

We used computational protein design to design a protein–DNA nanomaterial, whose co-assembly is purely driven by non-covalent interactions. Unlike assemblies that rely on chemical conjugations,

non-covalent self-assemblies can be tuned by altering the reaction conditions (for example, temperature, pH, salt concentration). Indeed, our protein–DNA nanostructures will not form in high concentrations of salt because the protein–DNA electrostatic interaction is shielded (see Extended Data Fig. 7). Recently, a co-assembling system for protein nanomaterials was developed in which two different proteins are required for assembly<sup>13</sup>. Co-assembling systems provide several advantages over single-component self-assemblies, including better control over the localization and timing of assembly, and greater functional and structural versatility of the assembly, as each component can confer unique attributes, especially in the case of hybrid materials such as those designed here. The protein–DNA nanostructures we designed could be further functionalized with the incorporation of engineered DNA structures, such as DNA origami or DNA aptamers. Furthermore, dualENH could be fused to peptide tags for antibody recognition or used for the specific attachment of organic and inorganic materials<sup>30</sup>. Finally, peptide self-assembly techniques<sup>7</sup> could be incorporated into the protein–DNA framework pre-



**Figure 4 | Co-crystal structure of protein–DNA complex.** **a**, dualENH forms a symmetric homodimer using helix 1 and helix 2 (green) as the protein–protein interface. Helix 3 (blue) binds to the dsDNA in the same way that wild-type ENH does. **b**, Two forms of dualENH are present in the co-crystal structure and occur in a molar ratio of 3:1 (green:cyan). **c**, **d**, Two forms of protein–DNA binding are observed in the co-crystal structure and occur in a

molar ratio of 1:1. Both have two dualENH homodimers bound on the designed 11-nucleotide motif TAATTTAATTT. In **c**, both of the dualENH dimers bind in the optimal motif (TAATTT) orientation, whereas in **d**, one of the dualENH dimers (right) binds in the suboptimal orientation (AAATTA, the reverse complementary sequence of the optimal motif). **e**, Slightly kinked nanowire structure found in the co-crystal structure.



sented here, and more sophisticated two-dimensional or three-dimensional protein–DNA co-assemblies with well-defined structures could be explored using other known protein–DNA-binding motifs. We anticipate that protein–DNA co-assembly will open up many new possibilities for advanced biomaterial designs.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 March; accepted 30 June 2015.**

**Published online 2 September 2015.**

- Sarikaya, M., Tamerler, C., Jen, A. K. Y., Schulten, K. & Baneyx, F. Molecular biomimetics: nanotechnology through biology. *Nature Mater.* **2**, 577–585 (2003).
- Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).
- Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
- Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
- Douglas, S. M. et al. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
- Zhang, S., Holmes, T., Lockshin, C. & Rich, A. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl Acad. Sci. USA* **90**, 3334–3338 (1993).
- Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nature Biotechnol.* **21**, 1171–1178 (2003).
- Fletcher, J. M. et al. Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
- Brodin, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nature Chem.* **4**, 375–382 (2012).
- Lai, Y. T., Cascio, D. & Yeates, T. O. Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129 (2012).
- King, N. P. et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
- Lanci, C. J. et al. Computational design of a protein crystal. *Proc. Natl Acad. Sci. USA* **109**, 7304–7309 (2012).
- King, N. P. et al. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
- Huang, P. S. et al. High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
- Thomson, A. R. et al. Computational design of water-soluble  $\alpha$ -helical barrels. *Science* **346**, 485–488 (2014).
- Chang, M., Yang, C. S. & Huang, D. M. Aptamer-conjugated DNA icosahedral nanoparticles as a carrier of doxorubicin for cancer therapy. *ACS Nano* **5**, 6156–6163 (2011).
- Amir, Y. et al. Universal computing by DNA origami robots in a living animal. *Nature Nanotechnol.* **9**, 353–357 (2014).
- Sanchez, C., Julian, B., Belleville, P. & Popall, M. Applications of hybrid organic–inorganic nanocomposites. *J. Mater. Chem.* **15**, 3559–3592 (2005).
- Grigoryan, G. et al. Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076 (2011).
- Niemeyer, C. M., Koehler, J. & Wuerdemann, C. DNA-directed assembly of bienzymic complexes from *in vivo* biotinylated NAD(P)H:FMN oxidoreductase and luciferase. *ChemBioChem* **3**, 242–245 (2002).
- Sano, T., Smith, C. L. & Cantor, C. R. Immuno-PCR: very sensitive antigen detection by means of specific antibody–DNA conjugates. *Science* **258**, 120–122 (1992).
- Wilner, O. I., Shimron, S., Weizmann, Y., Wang, Z. G. & Wilner, I. Self-assembly of enzymes on DNA scaffolds: en route to biocatalytic cascades and the synthesis of metallic nanowires. *Nano Lett.* **9**, 2040–2043 (2009).
- Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).
- Draganescu, A. & Tullius, T. D. The DNA binding specificity of engrailed homeodomain. *J. Mol. Biol.* **276**, 529–536 (1998).
- Guerrero, L., Smart, O. S., Woolley, G. A. & Allemann, R. K. Photocontrol of DNA binding specificity of a miniature engrailed homeodomain. *J. Am. Chem. Soc.* **127**, 15624–15629 (2005).
- Marshall, S. A., Morgan, C. S. & Mayo, S. L. Electrostatics significantly affect the stability of designed homeodomain variants. *J. Mol. Biol.* **316**, 189–199 (2002).
- Shah, P. S. et al. Full-sequence computational design and solution structure of a thermostable protein variant. *J. Mol. Biol.* **372**, 1–6 (2007).
- Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.-J. & Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1505072112> (2015).
- Mou, Y., Huang, P.-S., Thomas, L. M. & Mayo, S. L. Using molecular dynamics simulations as an aid in the prediction of domain swapping of computationally designed protein variants. *J. Mol. Biol.* **427**, 2697–2706 (2015).
- Whaley, S. R., English, D. S., Hu, E. L., Barbara, P. F. & Belcher, A. M. Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly. *Nature* **405**, 665–668 (2000).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This study was supported by the Defense Advanced Research Projects Agency Protein Design Processes Program, a National Security Science and Engineering Faculty Fellowship (NSSEFF N00244-09-1-0011, N00244-09-1-0082), and the Gordon and Betty Moore Foundation through grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative. We would like to acknowledge the Gordon and Betty Moore Foundation for support of the Molecular Observatory at Caltech, and the Department of Energy and National Institutes of Health for supporting the Stanford Synchrotron Radiation Lightsources. We thank J. Kaiser, J. Hoy and P. Nikolovski at the Caltech Molecular Observatory for assistance in crystal screening and crystallographic data collection. Y.M. thanks L.-C. Ho for her encouragement and literature research in the crystallographic work. Y.M. thanks T. J. Zwang for assistance with AFM measurements. Y.M. thanks X. Zhang and S. Yan for the useful discussion. We are grateful to J. Kostecki and M. Ary for assistance with the manuscript.

**Author Contributions** Y.M. designed and performed the experiments. Y.M. and J.-Y.Y. performed the optical microscope experiments. All authors wrote the manuscript.

**Author Information** The co-crystal structure of the protein–DNA complex has been deposited in the PDB under accession number 4QTR. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.M. ([mouung@caltech.edu](mailto:mouung@caltech.edu)) or S.L.M. ([steve@mayo.caltech.edu](mailto:steve@mayo.caltech.edu)).

## METHODS

**Protein docking and computational design.** The ENH crystal structure (PDB accession 1ENH)<sup>31</sup> was used as the scaffold for homodimerization. A symmetrical docking program based on a fast Fourier transform (FFT) algorithm was applied<sup>32</sup>. Using shape complementarity as the criterion, the top 200 docking models were selected and clustered into 11 groups based on structural similarity (r.m.s.d. values). These clusters were visually inspected and one model was chosen for homodimer design. In this homodimer model, the two helix 1s (one from each subunit) form a parallel helix–helix packing assembly and are separated by 10 Å, which is similar to the 9.8 Å separation found in naturally occurring coiled-coil dimers<sup>33</sup>. Homebuilt computational protein design software was used for symmetry-constrained homodimer designs. The following types of amino acids were allowed at the interfacial residues: A, D, E, F, H, I, K, L, N, Q, R, S, T, V, W and Y. The E28P mutation was identified by the molecular dynamics (MD) simulations as described elsewhere<sup>29</sup>. The Rosetta force field was used to rank sequence energies<sup>34</sup>. Sequence optimization was performed using an improved version of FASTER<sup>35</sup> using a rotamer library based on the backbone-dependent library described previously<sup>36</sup>. The sequences of dualENH and wild-type ENH are listed in Extended Data Table 1.

**Construct preparation, protein expression, and purification.** Oligonucleotides (Integrated DNA Technologies) containing ~20 bp overlapping segments were assembled via a modified Stemmer PCR method<sup>37</sup> using KOD Hot Start Polymerase (Novagen) to generate genes for wild-type ENH and dualENH. A His<sub>6</sub> tag and a Gly-Ser linker (GGSGG) were added to the carboxy terminus. Proteins were expressed using BL21 DE3 cells transformed by pET plasmids with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) in standard Luria Broth (LB) at 37 °C. Proteins were purified from supernatant of lysed cells using affinity chromatography (Ni<sup>2+</sup>-NTA, Qiagen) followed by size-exclusion chromatography (Superdex 75, Amersham Pharmacia). Expression of dualENH at 37 °C produced >10 mg of soluble protein per litre of *Escherichia coli* culture.

**Circular dichroism spectroscopy.** Circular dichroism (CD) studies were performed on an Aviv 62A DS spectropolarimeter equipped with a thermoelectric temperature controller. Samples were prepared in 100 mM sodium chloride and 20 mM sodium phosphate buffer at pH 7.5. Wavelength scans and temperature denaturations were carried out in cuvettes with a 0.1 cm path length at a protein concentration of ~10 μM. Three wavelength scans were performed at 25 °C for each sample and averaged. The thermal denaturation curve was collected at 222 nm from 0 °C to 99 °C, sampling every 1 °C separated by 2 min equilibration times (signal averaging time was 1 s). The refolding curve was collected after the thermal denaturation experiment using the same sample.

**Analytical ultracentrifugation.** dualENH was analysed on an XL-1 analytical ultracentrifuge equipped with an AnTi60 rotor (Beckman Coulter). Two-channel epon-filled centerpieces were used for the sedimentation velocity experiment. Cells were torqued to 130 lb inch and run at 60,000 r.p.m. Data were acquired at 280 nm at 20 °C in continuous mode. Data were first fit to the c(S) model (continuous distribution of sedimentation coefficient) and then converted to the c(M) model (continuous distribution of molecular weight). Time invariant noises and baseline offsets were corrected before fitting. A maximum entropy regularization confidence level of 0.95 was used in all the size distribution analyses.

**Polarization fluorescence assay.** Polarization fluorescence was measured at room temperature with a Synergy 2 (BioTek). All DNA oligonucleotides were purchased from Integrated DNA Technologies without further purification. The three probes have the following sequences: CGCAGTGTAAATTACCTCGAC (probe 1; bold indicates the ENH protein-binding site), CGCAGTGTACTTACCTCGAC (probe 2; bold indicates the ENH protein-binding site with a single mutation), and CAGGCAGCAGGTGTTGGACT (negative control). The 3' terminus of each probe was labelled with fluorescein. The dsDNA samples were prepared by mixing equimolar single-stranded DNA with its complementary sequence. The mixture was heated to 95 °C for 10 min and gradually cooled down to room temperature. dualENH was serially diluted in buffer containing 20 mM Tris-HCl and 100 mM NaCl at pH 8.0, except for the NaCl-dependent experiments in Extended Data Fig. 7a. Concentrations of all probes were kept at 25 nM. The total volume of each sample was kept at 200 μl. The measurements were taken after about a 10 min equilibration. The G factor was calibrated and kept at 0.87 for all samples.

**Fluorescence resonance energy transfer assay.** The fluorescence resonance energy transfer (FRET) emission spectrum was measured at room temperature with a Safire2 (Tecan) plate reader. The (TAA)<sub>5</sub> oligonucleotide was labelled at its 5' terminus with either Cy3 or Cy5. Preparations of dsDNA samples were made as described earlier. Samples for the FRET experiment were prepared by mixing

400 nM Cy3-(TAA)<sub>5</sub> and 600 nM Cy5-(TAA)<sub>5</sub> for a reference, and then by adding 4 μM dualENH to observe a FRET signal change. The buffer contained 20 mM Tris-HCl and 100 mM NaCl at pH 8.0.

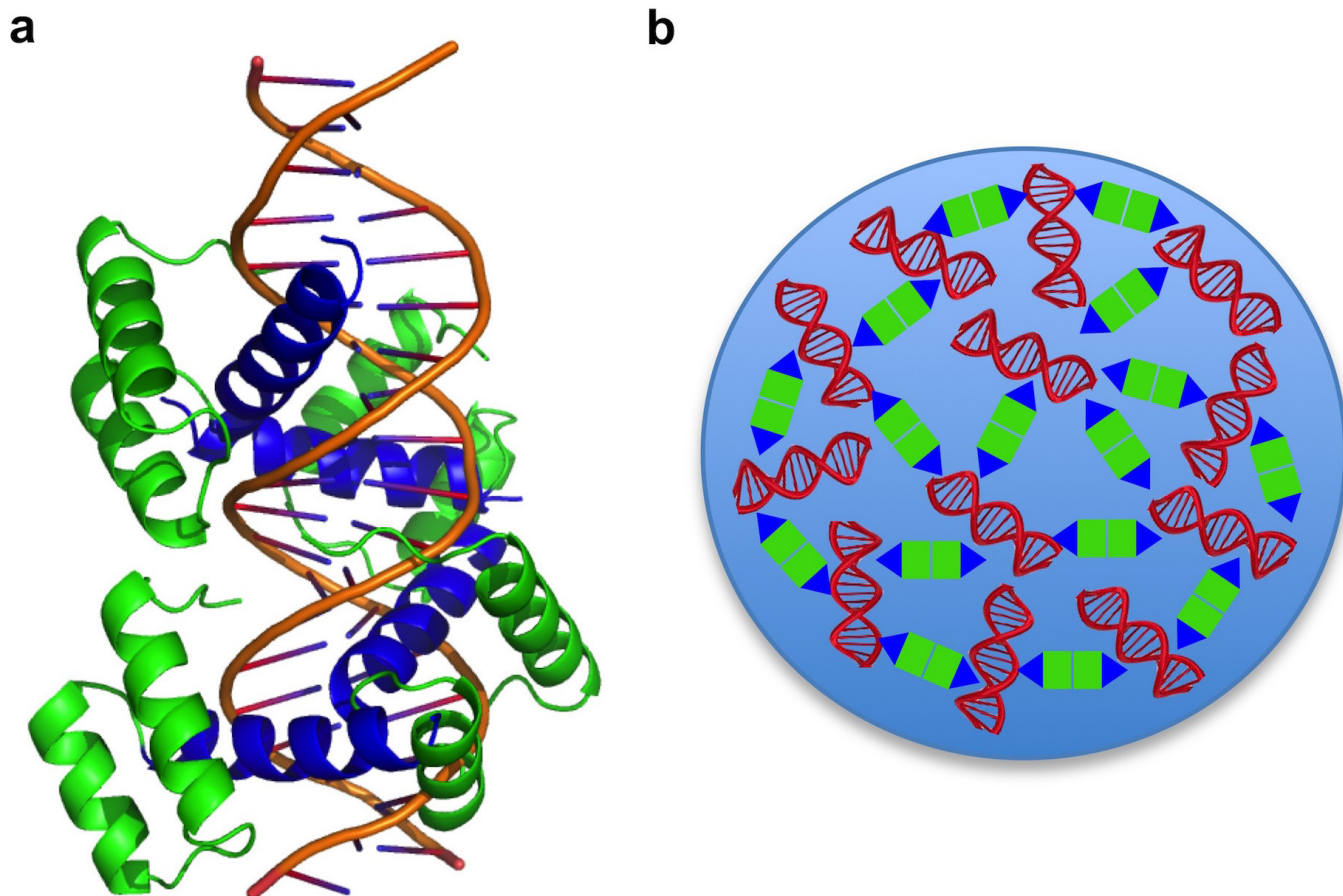
**Microscope imaging.** All imaging was performed at room temperature on a standard epifluorescence microscope (IX71, Olympus) equipped with bright-field and fluorescence modalities. The imaging objective was a ×40 NA 0.75 objective lens (UPLFLN 40X, Olympus). The (TAA)<sub>5</sub> oligonucleotides were labelled with Cy3 at their 5' terminus. The 11-nucleotide motif oligonucleotides (CGCAGTGTAAATTTAATTTCTCGAC; bold indicates motif 11) were labelled with fluorescein at their 5' terminus. The dsDNA samples were prepared as described earlier. All experiments were done in the following buffer: 20 mM Tris-HCl, 100 mM NaCl at pH 8.0, except that the NaCl concentration was increased to 150 mM in Extended Data Fig. 7b.

**Particle size distribution.** We used quantitative fluorescence microscopy to determine the particle size distribution of irregularly shaped protein–DNA particles. We assumed that, for a particle formation process with no specific dimensional preferences, the particle volume is approximately proportional to the 3/2-th power of the particle area, which can be directly estimated from the fluorescence image. We plotted particle brightness against particle volume for particles with roundness (minor axis/major axis) >0.7, which represents ~35% of the population and spans all particle sizes, and applied a linear regression to determine the relationship between the particle brightness and volume for each image (averaged R<sup>2</sup> = 0.93). Using this brightness versus volume relationship, we calculated the volume of every particle, including those smaller than the imaging resolution and those with low roundness, to obtain an overall particle size distribution.

**Atomic force microscopy.** Samples were deposited on a mica surface in a buffer containing 100 mM NaCl, 4 mM MgCl<sub>2</sub> and 20 mM TrisHCl at pH 8.0. After a 2 min incubation, the mica surface was washed with 3 ml pure water and air-dried. AFM images were taken using repulsive AC mode on an Asylum MFP-3D-bio imager, with an AFM tip spring constant of 3 N m<sup>-1</sup>. The scanning rate was 1 Hz.

**X-ray crystallography.** The dsDNA used for crystallization had forward and backward sequences as follows: GTGTAATTTAATTTCC and CGGAAATTAATTACA, respectively. An equimolar mixture of the forward and backward oligomers was heated to 95 °C for 10 min and gradually cooled down to room temperature. The dualENH (4.9 mM in 1.6 M sodium chloride and 20 mM MES buffer at pH 5.8) and the dsDNA (5.6 mM in 10 mM TrisHCl at pH 8.0) were mixed in equal volumes. Protein–DNA co-crystals were grown at room temperature in 0.2 M potassium thiocyanate and 20% w/v polyethylene glycol 3350 at pH 7.0 using hanging-drop diffusion. Diamond-like crystals appeared within 1–2 days. The crystals were soaked in 25% ethylene glycol cryoprotectant and flash frozen by liquid nitrogen. Diffraction data were collected with a wavelength of 1.03 Å at beamline 12-2 at Stanford Synchrotron Radiation Lightsource. The best diffraction data had a resolution of ~3.1 Å. Phases were obtained through molecular replacement using wild-type ENH–DNA co-crystal structure (PDB accession 3HDD)<sup>38</sup> as the searching model. Further refinement was done with PHENIX<sup>39</sup> and Coot<sup>40</sup>. The final refined coordinate has 97.7% and 2.3% backbone dihedral angles in the most favoured and additional allowed regions, respectively. The data statistics are listed in Extended Data Table 2.

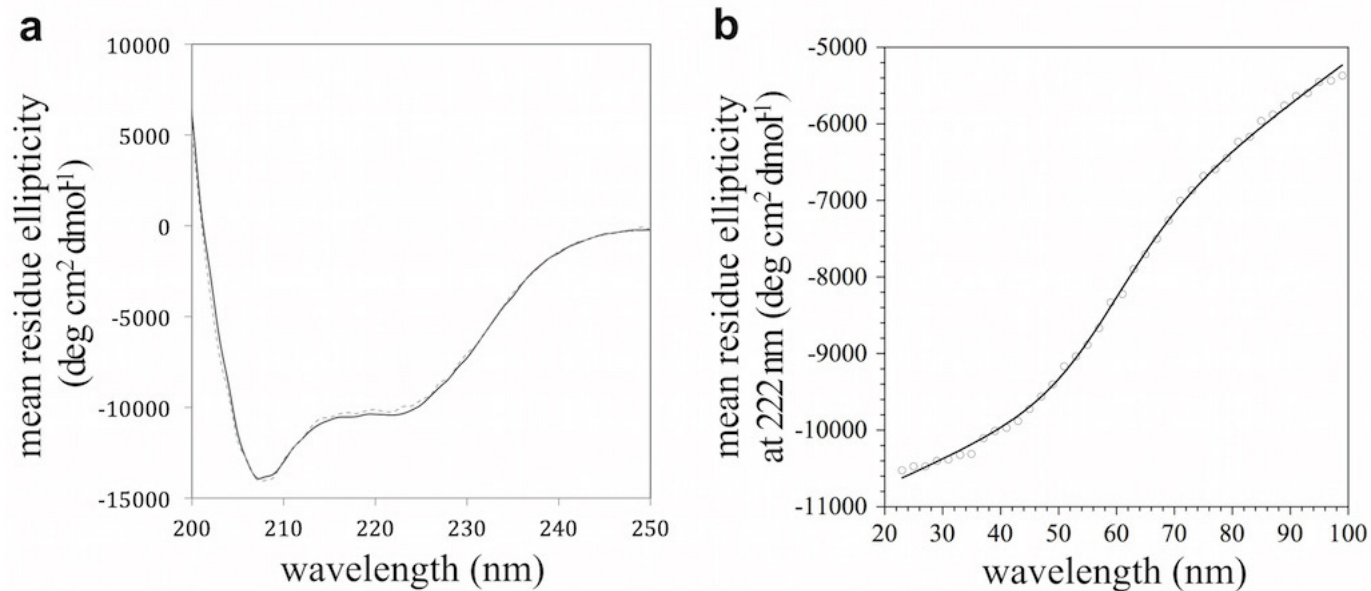
- Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L. & Pabo, C. O. Structural studies of the engrailed homeodomain. *Protein Sci.* **3**, 1779–1787 (1994).
- Huang, P.-S., Love, J. J. & Mayo, S. L. Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J. Comput. Chem.* **26**, 1222–1232 (2005).
- O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544 (1991).
- Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
- Allen, B. D. & Mayo, S. L. Dramatic performance enhancements for the FASTER optimization algorithm. *J. Comput. Chem.* **27**, 1071–1075 (2006).
- Dunbrack, R. L. Jr & Karplus, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol.* **230**, 543–574 (1993).
- Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxynucleotides. *Gene* **164**, 49–53 (1995).
- Fraenkel, E., Rould, M. A., Chambers, K. A. & Pabo, C. O. Engrailed homeodomain–DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.* **284**, 351–361 (1998).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).



**Extended Data Figure 1 | Design model of irregular bulk protein-DNA nanoparticle.** **a**, Four consecutive ENH-binding sites that each face in a different direction are engineered onto a dsDNA fragment. This dsDNA building block allows the protein-DNA assembly to occur in all three dimensions. Note that in this particular design, two neighbouring binding sites may

not be simultaneously occupied due to steric hindrance. **b**, Cartoon illustrating an irregular shaped nanoparticle formed by co-assembly of dualENH and the dsDNA shown in **a**. The DNA-binding domains of dualENH are shown as blue triangles, and the homodimerization domains are shown as green squares.

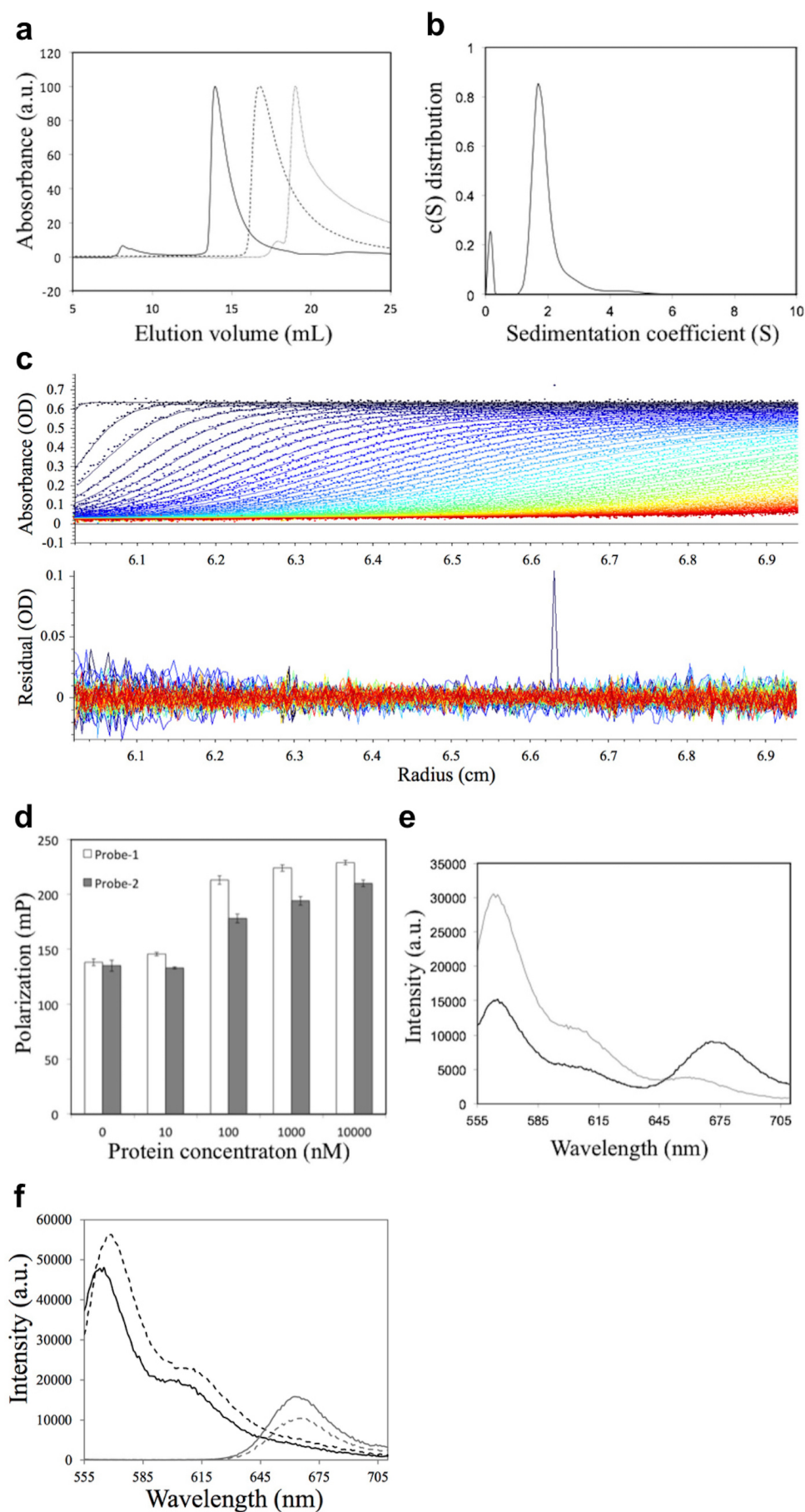




**Extended Data Figure 2 | Circular dichroism spectroscopy of dualENH.**

**a**, Circular dichroism spectrum of dualENH at room temperature. Solid line: before thermal denaturation; dashed line: after thermal denaturation. The overlapping of the two curves indicates that dualENH folds reversibly.

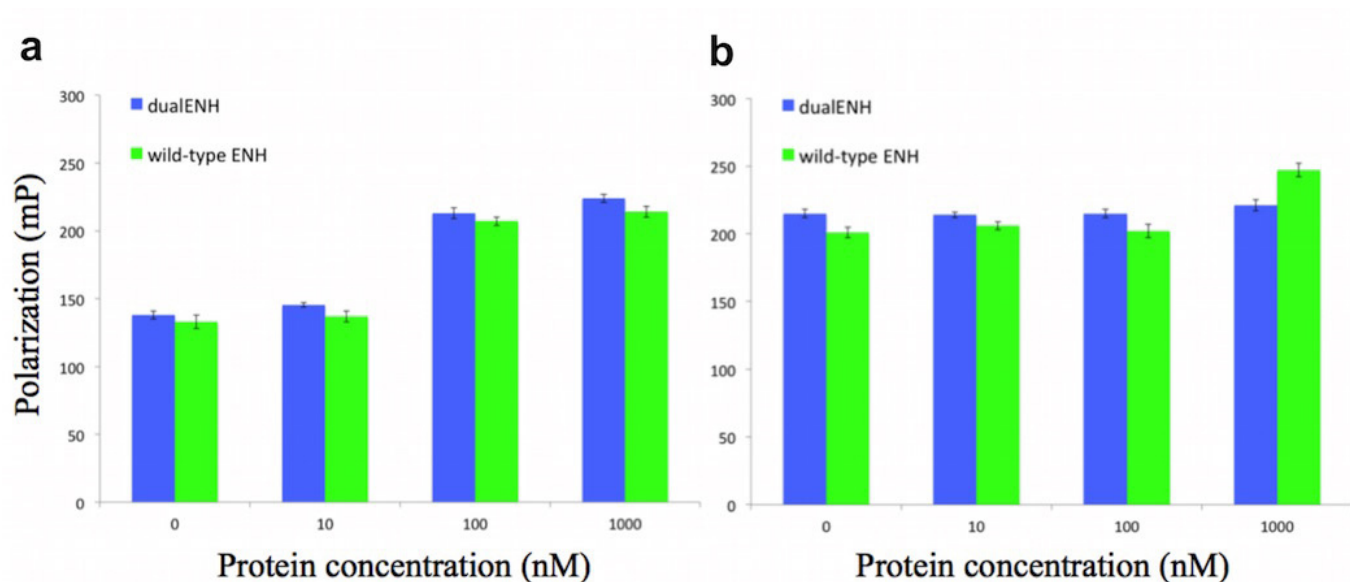
**b**, Thermal denaturation curve measured at 222 nm. Circles: experimental data; line: curve fit with two-state transition model. The melting temperature of dualENH was determined to be 59 °C.



**Extended Data Figure 3 | Biophysical characterization of dualENH.** **a**, Size-exclusion chromatography of dualENH with three different loading concentrations: solid line, 650  $\mu$ M; dashed line, 80  $\mu$ M; dotted line, 5  $\mu$ M. The highest signals were normalized to 100 for all curves. **b**,  $c(S)$  model fit from a sedimentation velocity experiment of 40  $\mu$ M dualENH. The major peak around  $S = 1.9$  corresponds to a molecular weight of 18.3 kDa, which is about twice that of monomeric dualENH (8.7 kDa). The spike at the left ( $S < 0.5$ ) may be due to impurities or artefacts from model fitting. **c**, Raw data and fitting residuals for the sedimentation velocity experiment in **b**. A total of 378 curves were used for fitting, but for visual clarity only one-fifth of the curves are shown. The top graph shows the raw data (dots) and the fitting curves; the bottom shows the residuals between the experimental data and the fit. The square root of variance of the fit is 0.00669. **d**, Fluorescence polarization experiment. Two dsDNA sequences labelled with fluorescein were used as probes to assay

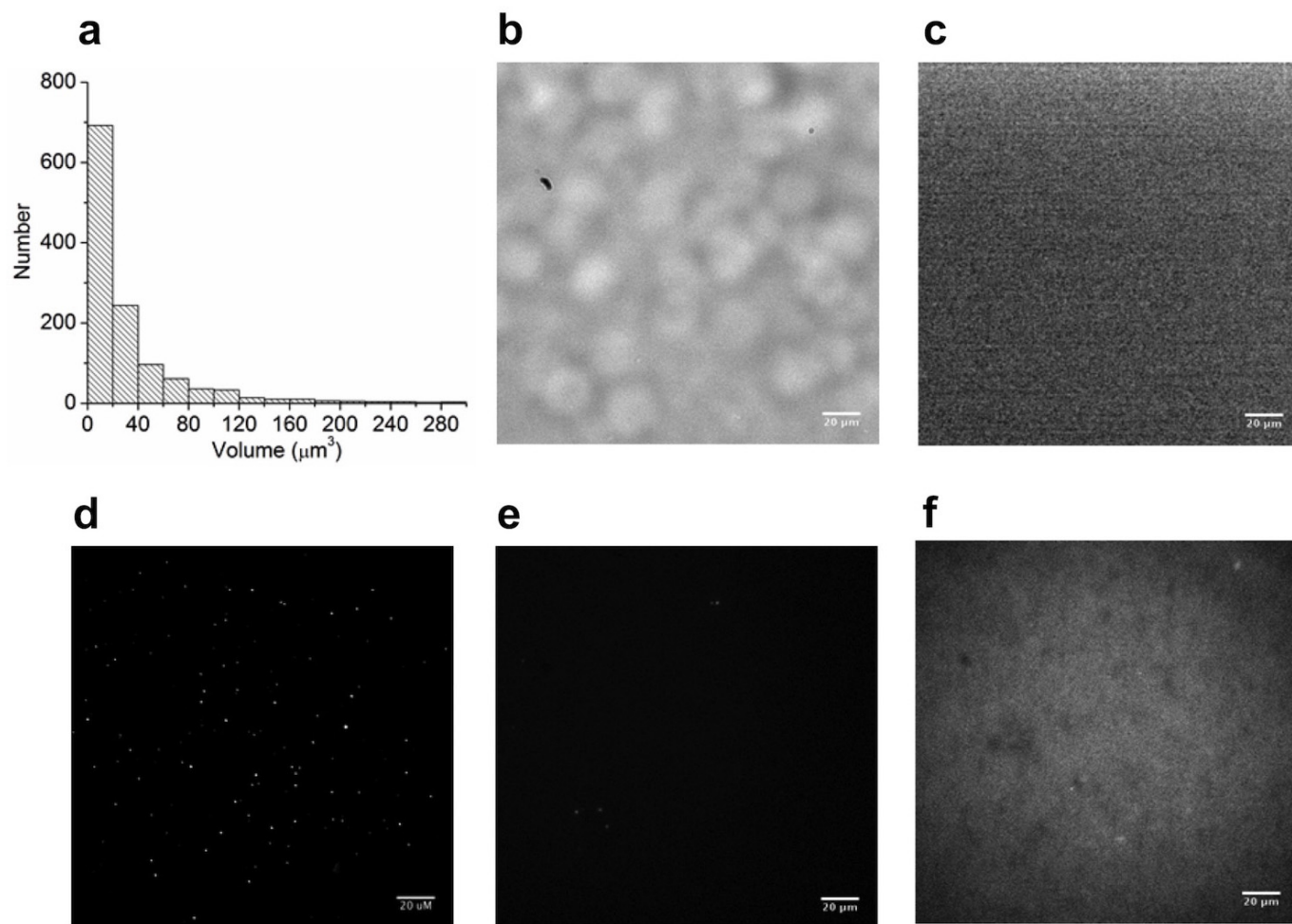
dualENH–DNA binding. Probe 1: 20-nucleotide dsDNA with the binding motif TAATTA; probe 2: same sequence as probe 1 but with a single-nucleotide mutation to the binding motif (TA[C]TTA). The concentration of dualENH was varied, while the concentration of the three probes remained constant (25 nM). Data are shown as mean  $\pm$  standard error of the mean (s.e.m.) for three replicates. **e**, FRET experiment showing that dualENH brings two dsDNA fragments within Förster distance. Fifteen-nucleotide dsDNA (TAA)<sub>5</sub> were labelled with either Cy3 or Cy5 to serve as the FRET donor or acceptor. Grey line: 400 nM Cy3-(TAA)<sub>5</sub> plus 600 nM Cy5-(TAA)<sub>5</sub>; black line: 400 nM Cy3-(TAA)<sub>5</sub> plus 600 nM Cy5-(TAA)<sub>5</sub> plus 4  $\mu$ M dualENH. **f**, Two control experiments for the FRET experiment in **e**. Black line: 400 nM Cy3-(TAA)<sub>5</sub>; black dashed line: 400 nM Cy3-(TAA)<sub>5</sub> plus 4  $\mu$ M dualENH; grey line: 600 nM Cy5-(TAA)<sub>5</sub>; grey dashed line: 600 nM Cy5-(TAA)<sub>5</sub> plus 4  $\mu$ M dualENH.





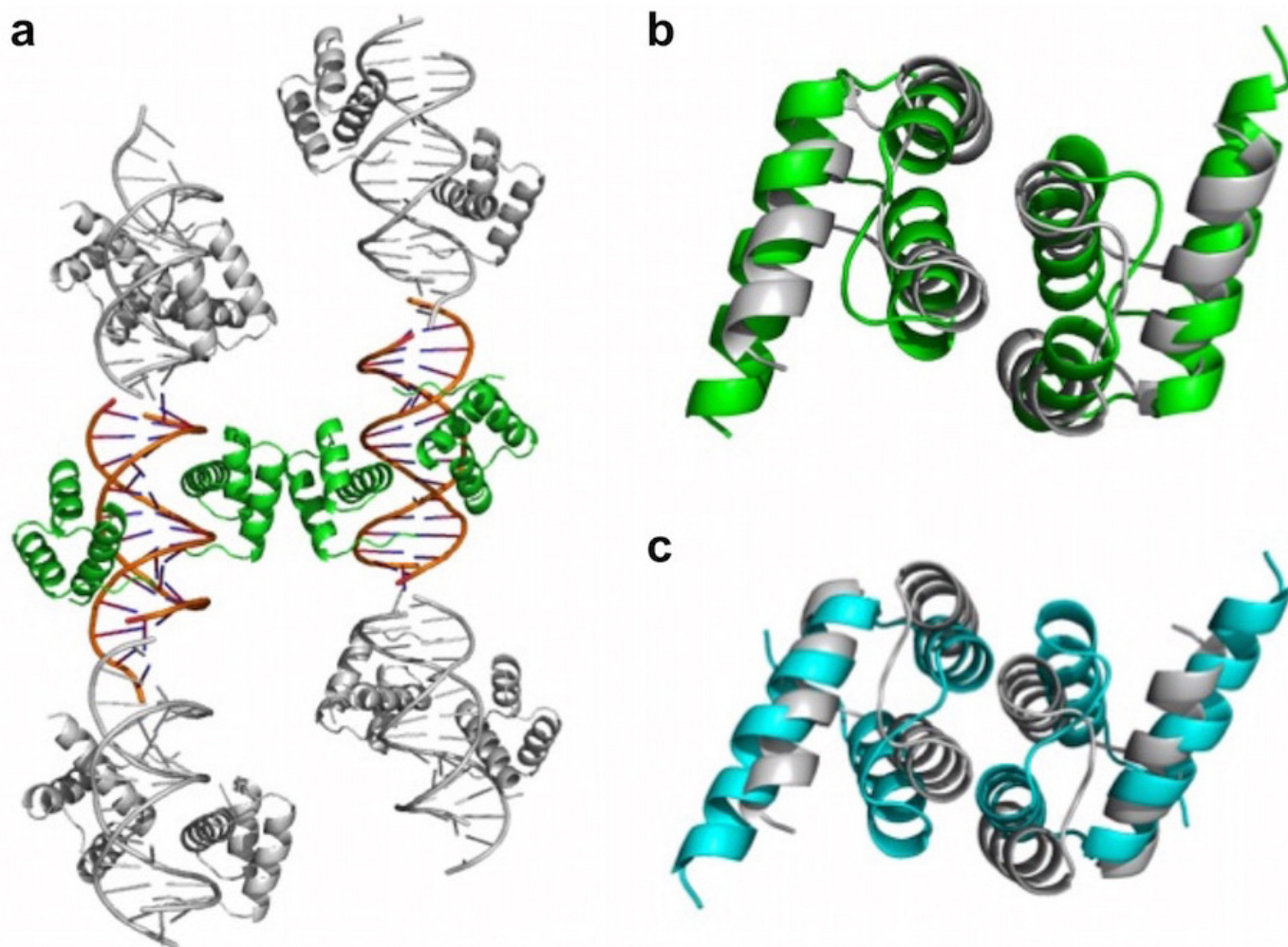
**Extended Data Figure 4 | Fluorescence polarization experiments with dualENH and wild-type ENH.** **a**, Fluorescence polarization experiment with probe 1, which contains the binding motif TAATTA of ENH. The concentration of dualENH or wild-type ENH was varied, while the concentration of probe 1 remained constant (25 nM). Both dualENH and wild-type ENH show saturated binding when the protein concentration is at or

above 100 nM. Data are shown as mean  $\pm$  s.e.m. for three replicates. **b**, Same experiments as **a**, except that a probe without any TAATNN binding motif was used. Note that probe 1 used in **a** has a lower fluorescence intensity and polarization than the probe used in **b**, probably due to partial quenching by a guanine nucleotide on the strand opposite the fluorescein label. Data are shown as mean  $\pm$  s.e.m. for three replicates.



**Extended Data Figure 5 | Microscope imaging experiments.** **a**, The size distribution of the irregular protein–DNA particles formed by 5  $\mu\text{M}$  dualENH mixed with 2  $\mu\text{M}$  Cy3-(TAA)<sub>5</sub>. **b**, Bright-field microscopy image of 5  $\mu\text{M}$  dualENH mixed with 2  $\mu\text{M}$  Cy3-(TAA)<sub>5</sub>. A dust particle (top left) is evident, indicating that the focal plane is correct. **c**, Fluorescence microscopy image of 2  $\mu\text{M}$  Cy3-(TAA)<sub>5</sub> alone. **d**, Fluorescence microscopy image of particles formed with 500 nM dualENH mixed with 200 nM Cy3-(TAA)<sub>5</sub>.

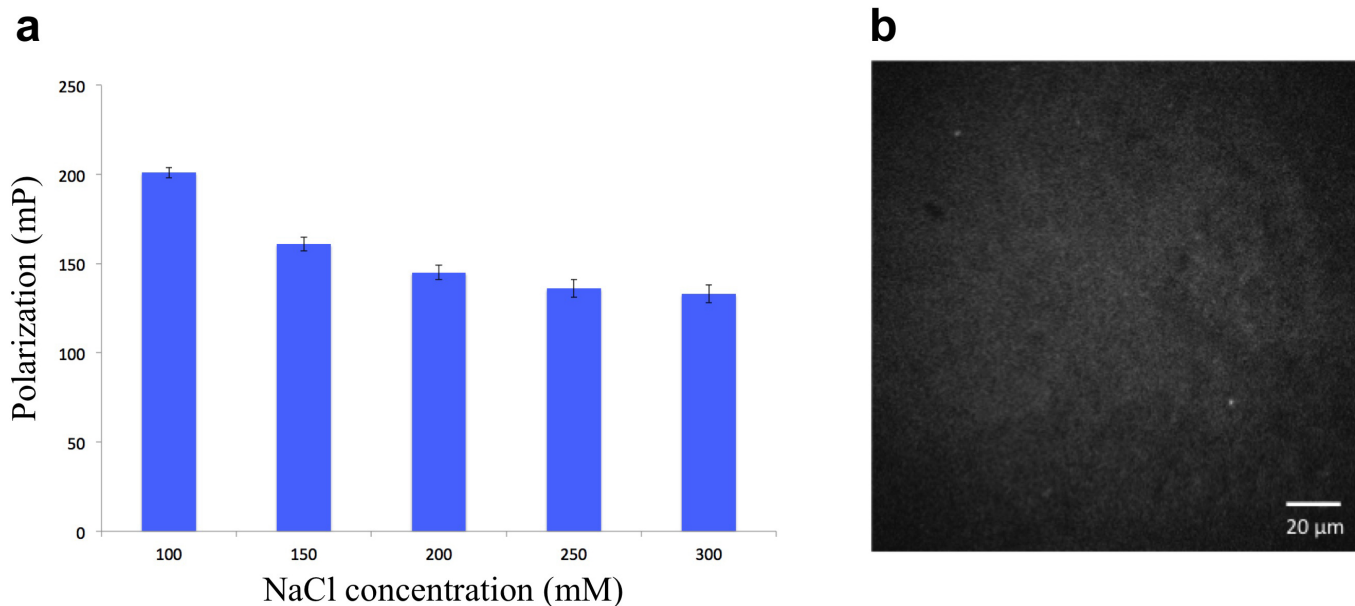
**e**, Fluorescence microscopy image of particles formed with 200 nM dualENH mixed with 100 nM Cy3-(TAA)<sub>5</sub>. **f**, Fluorescence microscopy image of particle inhibition experiments. A small amount (5 nM) of single-binding-site dsDNA (containing only one TAATTA motif) was pre-mixed with 500 nM dualENH, then 200 nM Cy3-(TAA)<sub>5</sub> was added. The illumination/camera sensitivity was enhanced to confirm that particle formation is nearly completely absent under these conditions.



**Extended Data Figure 6 | Co-crystal structure of the protein–DNA complex.** **a**, Structures in the asymmetric unit cell are shown in colour, and the end-to-end packing of neighbouring DNA molecules and their bound proteins are shown in grey. **b**, **c**, The dualENH homodimer observed in the co-crystal structure (green or cyan) is superimposed with the design model (grey). The backbone r.m.s.d. compared to the design model is 3.8 Å (green) and 3.9 Å (cyan), respectively. When only one subunit is aligned between the

more predominant configuration (green) and the design model, the angular displacement between the other subunits is  $\sim 45^\circ$  with about 3 Å translational displacement. The less predominant configuration has a lower angular displacement,  $\sim 20^\circ$ , but a larger translational displacement,  $\sim 8$  Å. The calculated energies for the design model and the two crystallographic dimers are  $-155.2$ ,  $-140.3$  (green) and  $-131.2$  (cyan) Rosetta energy units, respectively.






**Extended Data Figure 7 | dualENH–DNA binding and nanostructure formation are inhibited at high salt concentrations.** **a**, Fluorescence polarization experiments of dualENH and probe 1 at various NaCl concentrations. Probe 1 and dualENH were mixed in buffers with different NaCl concentrations and fluorescence polarization values were recorded. dualENH–DNA binding dropped greatly from 100 mM to 150 mM NaCl, and was

completely absent at 300 mM NaCl. Data are shown as mean  $\pm$  s.e.m. for three replicates. **b**, Fluorescence microscopy image of particle experiment at 150 mM salt concentration. The sample was prepared by mixing 500 nM dualENH and 200 nM Cy3-(TAA)<sub>5</sub> in 20 mM Tris-HCl buffer with 150 mM NaCl. The illumination/camera sensitivity was enhanced to confirm that particle formation is nearly completely absent under these conditions.

Extended Data Table 1 | Sequences of wild-type ENH and dualENH

	
ENH (WT)	EKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSELGLNEAQIKIWFQNKRAKIKKST
dualENH	-----E--KKA-DLA-YFD---PEW-RY--QR-----

The three 'coils' at the top show the location of the three helices in the wild-type (WT) fold based on PDB structure 1ENH. The homodimerization domain is located on the first and second helices, and the DNA-binding domain is located on the third helix.

Extended Data Table 2 | Co-crystal structure statistics for dualENH complexed with dsDNA containing motif 11 (PDB accession 4QTR)

	4QTR
<b>Data collection</b>	
Space group	P4222
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	90.1, 90.1, 158.9
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0
Resolution (Å)	39-3.1 (3.2-3.1)*
<i>R</i> <sub>merge</sub>	0.043 (1.456)
<i>I</i> / $\sigma$ <i>I</i>	28.4 (2.7)
Completeness (%)	99.8 (99.8)
Redundancy	12.7 (13.3)
<b>Refinement</b>	
Resolution (Å)	36-3.2
No. reflections	14,287
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	26/32
No. atoms	
Protein/DNA	2903
Ligand/ion	0
Water	0
B-factors	
Protein/DNA	174.5
Ligand/ion	n/a
Water	n/a
R.m.s deviations	
Bond lengths (Å)	0.012
Bond angles (°)	1.32

One crystal was used for this coordinate.

\*Highest-resolution shell is shown in parentheses.



# A marine biogenic source of atmospheric ice-nucleating particles

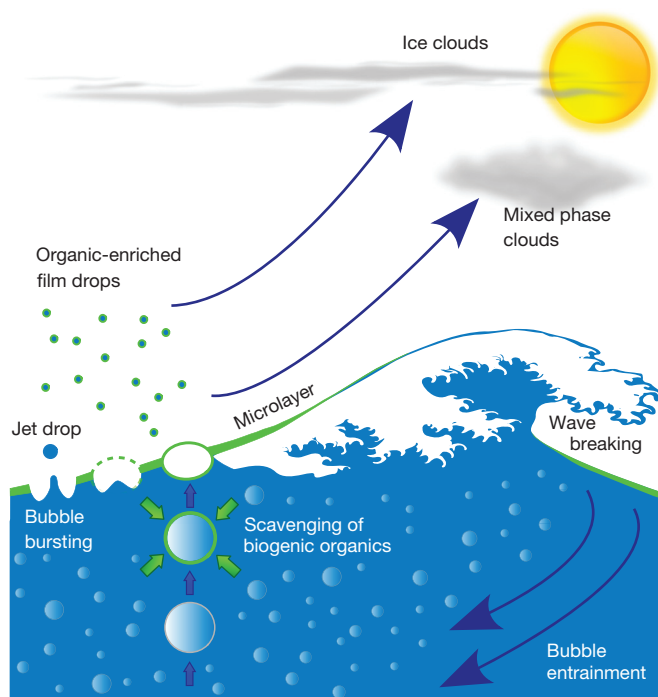
Theodore W. Wilson<sup>1\*</sup>, Luis A. Ladino<sup>2\*</sup>, Peter A. Alpert<sup>3</sup>, Mark N. Breckels<sup>4</sup>, Ian M. Brooks<sup>1</sup>, Jo Browne<sup>1</sup>, Susannah M. Burrows<sup>5</sup>, Kenneth S. Carslaw<sup>1</sup>, J. Alex Huffman<sup>6</sup>, Christopher Judd<sup>1</sup>, Wendy P. Kilhau<sup>7</sup>, Ryan H. Mason<sup>8</sup>, Gordon McFiggans<sup>9</sup>, Lisa A. Miller<sup>10</sup>, Juan J. Nájera<sup>9</sup>, Elena Polishchuk<sup>8</sup>, Stuart Rae<sup>9</sup>, Corinne L. Schiller<sup>11</sup>, Meng Si<sup>8</sup>, Jesús Vergara Temprado<sup>1</sup>, Thomas F. Whale<sup>1</sup>, Jenny P. S. Wong<sup>2</sup>, Oliver Wurl<sup>12†</sup>, Jacqueline D. Yakobi-Hancock<sup>2</sup>, Jonathan P. D. Abbatt<sup>2</sup>, Josephine Y. Aller<sup>7</sup>, Allan K. Bertram<sup>8</sup>, Daniel A. Knopf<sup>3</sup> & Benjamin J. Murray<sup>1</sup>

The amount of ice present in clouds can affect cloud lifetime, precipitation and radiative properties<sup>1,2</sup>. The formation of ice in clouds is facilitated by the presence of airborne ice-nucleating particles<sup>1,2</sup>. Sea spray is one of the major global sources of atmospheric particles, but it is unclear to what extent these particles are capable of nucleating ice<sup>3–11</sup>. Sea-spray aerosol contains large amounts of organic material that is ejected into the atmosphere during bubble bursting at the organically enriched sea–air interface or sea surface microlayer<sup>12–19</sup>. Here we show that organic material in the sea surface microlayer nucleates ice under conditions relevant for mixed-phase cloud and high-altitude ice cloud formation. The ice-nucleating material is probably biogenic and less than approximately 0.2 micrometres in size. We find that exudates separated from cells of the marine diatom *Thalassiosira pseudonana* nucleate ice, and propose that organic material associated with phytoplankton cell exudates is a likely candidate for the observed ice-nucleating ability of the microlayer samples. Global model simulations of marine organic aerosol, in combination with our measurements, suggest that marine organic material may be an important source of ice-nucleating particles in remote marine environments such as the Southern Ocean, North Pacific Ocean and North Atlantic Ocean.

Atmospheric ice-nucleating particles (INPs) allow ice to nucleate heterogeneously at higher temperatures or lower relative humidity than is typical for homogeneous ice nucleation. Heterogeneous ice nucleation proceeds via different pathways depending on temperature and humidity. In low-altitude mixed-phase clouds, INPs are commonly immersed in supercooled liquid droplets and freezing can occur on them at temperatures between  $-36^{\circ}\text{C}$  and  $0^{\circ}\text{C}$  (ref. 2). At higher altitudes and lower temperatures (less than  $-36^{\circ}\text{C}$ , the conditions under which cirrus clouds form), nucleation occurs below water saturation, proceeding by homogeneous, deposition or immersion-in-solution nucleation<sup>1</sup>. Understanding the sources of atmospheric INPs is important because they affect cloud lifetime, cloud albedo and precipitation<sup>1,2</sup>. Recent modelling work has shown that the ocean is potentially an important source of biogenic atmospheric INPs, particularly in remote, high-latitude regions<sup>9,10</sup>. However, it has never been directly shown that there is a source of atmospheric INPs associated with organic material found in marine waters or sea-spray aerosol.

Organic material makes up a substantial fraction of submicrometre sea-spray aerosol and it is estimated that  $10 \pm 5 \text{ Tgyr}^{-1}$  of primary

organic submicrometre aerosol is emitted from marine sources globally<sup>12</sup>. Rising bubbles scavenge surface active organic material from the water column at their interfaces and this process facilitates the formation of the organic enriched sea–air interface known as the sea surface microlayer (SML). This organic material is ejected into the atmosphere during bubble bursting, resulting in sea-spray aerosol



**Figure 1 | Sea-spray aerosol particles enriched in organic material are generated when bubbles burst at the air–sea interface.** Surface active organic material of biological origin is scavenged at the interfaces of bubbles as they rise through the water column. This process enriches the air–sea interface with surface active organic material forming the SML (green layers). The organic material is ejected on bubble bursting with resulting submicrometre film drops being enriched with organic material compared with larger jet drops. We show that the biogenic organic material in the SML is probably an important source of atmospheric INPs that could influence cloud properties.

<sup>1</sup>School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK. <sup>2</sup>Department of Chemistry, University of Toronto, 80 St George Street, Toronto, Ontario M5S 3H6, Canada.

<sup>3</sup>Institute for Terrestrial and Planetary Atmospheres, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York 11794-5000, USA. <sup>4</sup>School of Biological Sciences,

University of Essex, Colchester CO4 3SQ, UK. <sup>5</sup>Atmospheric Science and Global Change Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, Washington 99354, USA.

<sup>6</sup>Department of Chemistry and Biochemistry, University of Denver, Denver, Colorado 80208, USA. <sup>7</sup>School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York

11794-5000, USA. <sup>8</sup>Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada. <sup>9</sup>School of Earth, Atmospheric and Environmental Sciences,

University of Manchester, Manchester M13 9PL, UK. <sup>10</sup>Institute of Ocean Sciences, Fisheries and Oceans Canada, Sidney, British Columbia V8L 4B2, Canada. <sup>11</sup>Air Quality Science Unit, Environment

Canada, Vancouver, British Columbia V6C 3S5, Canada. <sup>12</sup>Leibniz Institute for Baltic Sea Research Warnemünde, Department of Biological Oceanography, Seestraße 15, 18119 Rostock, Germany.

<sup>†</sup>Present address: Carl-von-Ossietzky University Oldenburg, Institute of Chemistry and Biology of the Marine Environment, AG Meeresoberflächen, Emstrasse 20, 26382 Wilhelmshaven, Germany.

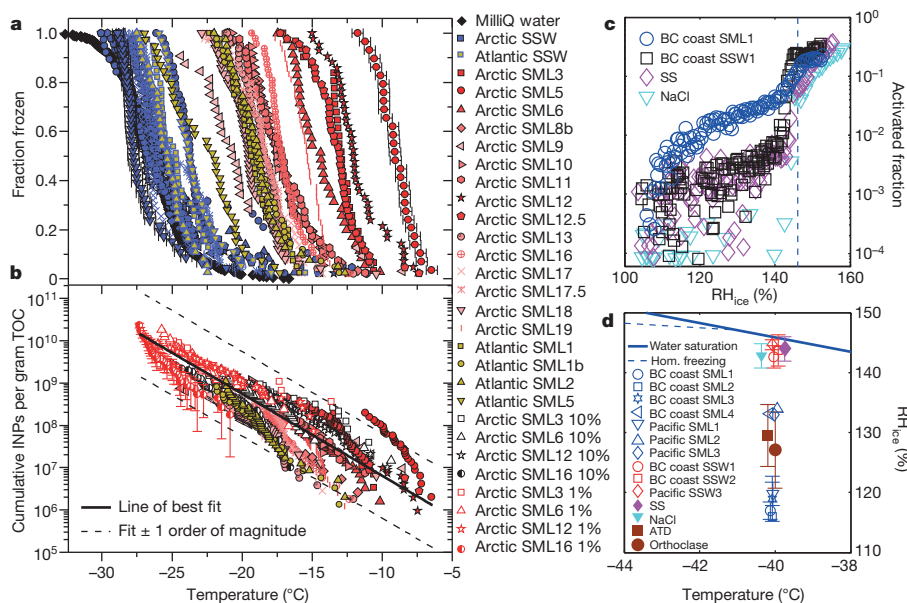
\*These authors contributed equally to this work.

containing similar organic material to that of the microlayer<sup>12–19</sup> (Fig. 1). Measurements of atmospheric INP concentrations in the remote oceans indicate that there may be a marine source of INPs linked to marine biology<sup>3–5</sup>, and modelling work indicates that this may be important in mixed-phase clouds<sup>9,10</sup>. There is also evidence to suggest that ice nucleation in cirrus clouds over ocean regions is influenced by sea-spray aerosol, with about 25% of the heterogeneously nucleated ice particle residuals identified as sea salt<sup>20</sup>. Despite these indications that there is a marine source of atmospheric INPs, the SML has not previously been analysed for the presence of material capable of nucleating ice. To determine whether there is a source of atmospheric INPs in the microlayer, we examined the ice-nucleating properties of microlayer samples collected in the Arctic (July–August 2013), Atlantic (May–June 2014), North Pacific (June 2013), and in coastal locations off British Columbia, Canada (August 2013) (Extended Data Fig. 1).

First we present experiments relevant to mixed-phase clouds, in which 1  $\mu$ l droplets of microlayer samples from the Arctic and Atlantic Oceans were placed on a cold stage immediately after sampling, and cooled until frozen. The fraction of droplets that froze as a function of temperature, corrected for freezing depression caused by salts, is shown in Fig. 2a. The microlayer droplets consistently froze at higher temperatures than droplets of subsurface water (SSW) collected at depths of between 2 to 5 m at the same locations (Extended Data Table 1). Filtration and re-testing of the microlayer samples showed that most material that nucleated ice was between approximately 0.2 and 0.02  $\mu$ m in size (Extended Data Fig. 2a). Material of this size has the potential to be lofted into the atmosphere through bubble-bursting processes, forming atmospheric INPs that are internally mixed with sea salt and other organics. To estimate atmospheric INP concentrations associated with marine organics, we determined the number of these INPs present per mass of organic carbon (Fig. 2b; see Methods); this result is used in the modelling section of this paper.

Experiments were also conducted under conditions relevant to cirrus clouds using microlayer and SSW samples from the North Pacific and the British Columbia coast. The activity of the collected samples was tested at  $-40^{\circ}\text{C}$  and compared with results from experiments with commercial sea salt and NaCl particles. Figure 2c shows example activation curves for aerosolized, dried and size-selected (200 nm diameter) particles. The SSW activation curves are very similar to those of sea salt and NaCl, with all showing sharp increases above 143% relative humidity with respect to ice ( $\text{RH}_{\text{ice}}$ ). This is consistent with homogeneous ice nucleation of solution droplets and suggests that crystalline salt particles did not contribute markedly to ice nucleation events observed at low  $\text{RH}_{\text{ice}}$ . In contrast, aerosol particles derived from microlayer samples all showed ice formation above the background level at lower  $\text{RH}_{\text{ice}}$ . The ice nucleation onset ( $\text{RH}_{\text{ice}}$  at which 1% of the aerosol particles were activated; Fig. 2d) varied between 115% and 133%  $\text{RH}_{\text{ice}}$ , which is comparable with efficient deposition mode INPs, such as Arizona test dust (ATD) and feldspar dust (orthoclase) particles of the same size<sup>1,21</sup>. Filtration of the SML samples through filters with nominal pore sizes of 0.2  $\mu$ m increased the ice nucleation onset by about  $12\text{--}16 \pm 4\%$   $\text{RH}_{\text{ice}}$ , indicating that some ice active material larger than 0.2  $\mu$ m was present (Extended Data Fig. 2b). However, the onsets for the 0.2- $\mu$ m-filtered samples remained well below the homogeneous threshold, indicating that there is a population of smaller INPs in the British Columbia and Pacific microlayer samples, consistent with the Arctic and Atlantic data. The cumulative number of ice nucleation sites per surface area as a function of  $\text{RH}_{\text{ice}}$ ,  $n_s(\text{RH}_{\text{ice}})$ , was greater for the microlayer samples than that found for ATD, kaolinite and feldspar mineral dusts<sup>21,22</sup> (Extended Data Fig. 3; see refs 1, 2 for more information on  $n_s$ ).

We emphasize that owing to the different nucleation processes involved, ice nucleation under mixed-phase cloud and cirrus conditions cannot be quantitatively compared. However, our results do clearly show that the microlayers at all the sampling locations were enriched in INPs compared with the SSW at the same locations. Here



**Figure 2 | Ice nucleation by material in the SML.** **a**, Immersion mode fraction frozen curves for 1  $\mu$ l Arctic and Atlantic SML and SSW droplets determined using the microlitre nucleation by immersed particle instrument ( $\mu$ l-NIPI), with example temperature uncertainties included. **b**, The cumulative number of INPs per gram of total organic carbon (TOC). Selected samples were diluted with ultrapure (MilliQ) water to 10% and 1% of initial concentration. Uncertainties are included where error bars are larger than data points (see Methods for details). Equation for fit to data is  $\text{INPs per gram TOC} = \exp[11.2186 - (0.4459 \times T)]$ ; note that temperature,  $T$ , is in  $^{\circ}\text{C}$ . **c**, Ice nucleation by British Columbia (BC) coast

and North Pacific samples under cirrus conditions. Example University of Toronto continuous flow diffusion chamber (UT-CFDC) activation curves under cirrus conditions (with background counts subtracted) for sea salt (SS), NaCl, British Columbia coast SML and SSW samples at  $-40^{\circ}\text{C}$ . **d**, Ice nucleation onset  $\text{RH}_{\text{ice}}$  for Pacific and British Columbia coast SML, SSW, sea salt and NaCl aerosol particles. For comparison, onsets for ATD and K-feldspar (orthoclase)<sup>21</sup> are shown. Ice-onset error bars represent one standard deviation based on three to four replicates. The solid line represents the water saturation line, and the dashed line is the theoretical homogeneous freezing threshold<sup>31</sup>.

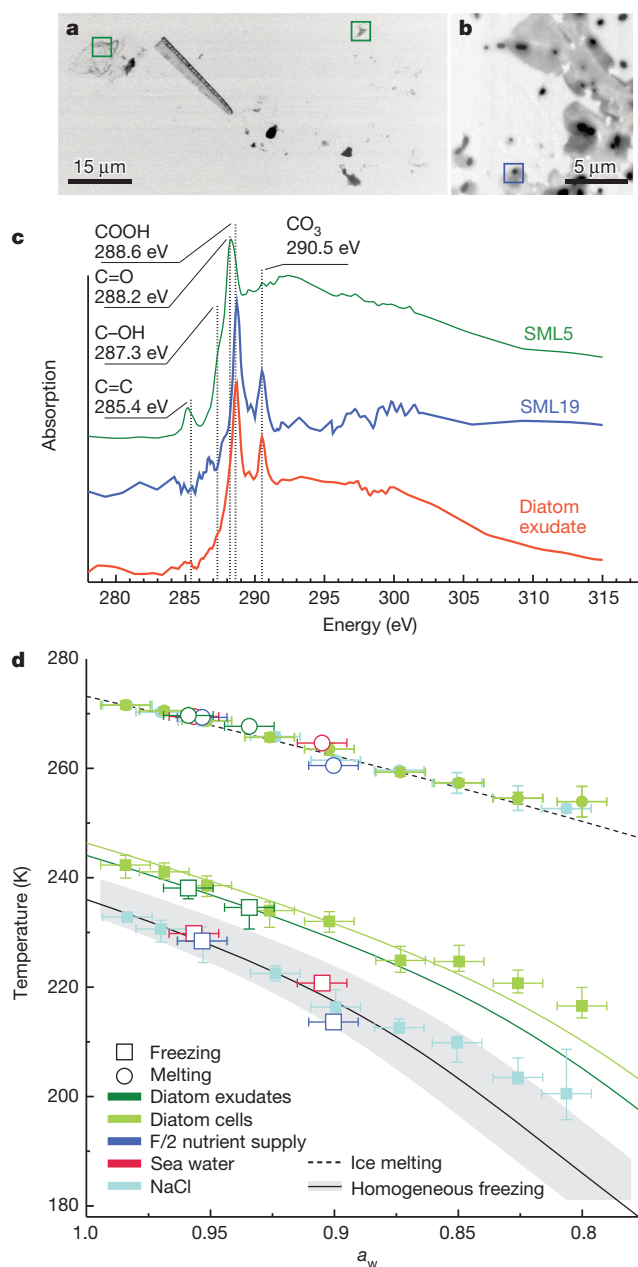
we present experiments designed to reveal likely candidates for the source of the ice-nucleating activity.

Certain proteins have been identified as being highly ice active; however, heat denatures proteins, causing a reduction in activity, which is not the case for known inorganic INPs<sup>2</sup>. To test whether similar thermally labile material could be responsible for the observed ice nucleation, microlayer samples were heated at temperatures up to 100 °C and retested for activity. Heating of Arctic and Atlantic samples resulted in a reduction of the ice nucleation activity with freezing shifting to lower temperatures (Extended Data Fig. 2c). Similarly, the onset  $RH_{ice}$  of a British Columbia coast sample increased by  $6 \pm 4\%$   $RH_{ice}$ , whereas that of a Pacific sample was within uncertainty of the unheated sample (Extended Data Fig. 2b). This might be consistent with the presence of inorganic or other non-thermally labile INPs. The marked reduction in activity in samples from three out of four locations is consistent with the presence of thermally labile biological INPs.

The filtration tests on the sampled microlayers (mentioned earlier) show that there is a considerable population of INPs that pass through 0.2  $\mu\text{m}$  filters (Extended Data Fig. 2a, b). The ice active materials are therefore probably smaller biological particles, for example, ultramicrobacteria, viruses or extracellular material from phytoplankton or bacteria (exudate). Additionally, no correlation was found between freezing temperature and bacterial cell counts in the Arctic microlayer (Extended Data Fig. 4), which suggests that whole bacterial cells were not responsible for the observed ice nucleation. Given that terrestrial biological systems such as pollens<sup>23,24</sup> and fungi<sup>25,26</sup> have been found to produce nanoscale or ‘macromolecular’ INPs unconnected with whole cells, we considered the possibility that marine INPs are associated with exudates from phytoplankton or other marine microorganisms. This hypothesis is not only supported by the filtration tests, but by a tentative correlation between the North Pacific microlayer sample ice activation onsets with both the dissolved organic carbon concentration (DOC;  $>0.2 \mu\text{m}$ ) and polysaccharide-rich transparent exopolymer particles (TEPs), which are associated with phytoplankton exudates (Extended Data Fig. 5).

Qualitative compositional analysis of two Arctic samples using scanning transmission X-ray microscopy coupled with near edge absorption fine structure spectroscopy (STXM/NEXAFS; Fig. 3a–c) indicates the presence of both diatom cell wall and exudate compounds (see Methods). Spectra of exudates from the ubiquitous marine diatom *T. pseudonana*<sup>27</sup> share absorption features with the microlayer samples. These data are in keeping with studies showing that diatom exudates are present in microlayer samples<sup>18</sup> and consistent with the fact that diatoms are the dominant phytoplanktonic group in polar regions<sup>28</sup>.

Diatom cells and fragments have been shown to nucleate ice heterogeneously<sup>6</sup> but, as demonstrated, whole cells are not solely responsible for the ice nucleation activity we observe in the microlayer samples. Here we investigate whether exudates separated from *T. pseudonana* diatom cells can nucleate ice heterogeneously. The ice nucleation efficiency of exudate from an axenic unialgal culture of *T. pseudonana* filtered through a 0.1  $\mu\text{m}$  filter was measured as a function of temperature and water activity ( $a_w$ ) in nanolitre volume droplets. Exudate freezing temperatures were found to be similar to those of washed diatom cells in the absence of exudate material and approximately 9–13 °C warmer than observed homogeneous freezing temperatures of 0.2- $\mu\text{m}$ -filtered and autoclaved Atlantic sea water collected 100 km offshore of Long Island, New York (the same water used to culture diatoms) with and without added nutrients (Fig. 3d; freezing curves are shown in Extended Data Fig. 6). While the freezing temperatures shown in Fig. 3d are not directly comparable to the microlayer droplet experiments in Fig. 2a, as they used much smaller droplets, they do show that material associated with exudates can nucleate ice.



**Figure 3 | Spectroscopic analysis of Arctic SML samples and freezing experiments with diatom exudate.** **a, b**, X-ray images of Arctic SML5 (**a**) and SML19 (**b**). Example locations at which spectra were acquired are indicated by green and blue boxes. **c**, X-ray absorption spectra of organic material in SML5, SML19 and exudates from the diatom *T. pseudonana*. **d**, Freezing and melting temperatures, collected using the water-activity-controlled immersion freezing experiment (WACIFE), as a function of water activity ( $a_w$ ) for nanolitre volume droplets containing *T. pseudonana* exudates, and filtered and autoclaved natural sea water with and without f/2 nutrients. Heterogeneous ice nucleation temperatures in the presence of diatom cells and homogeneous ice nucleation of aqueous NaCl droplets are shown for comparison<sup>6,32</sup>. Vertical error bars represent 10th and 90th percentiles of about 300 individual freezing events. Horizontal error bars indicate the uncertainty in  $a_w$  of  $\pm 0.01$ .

Given the ice nucleation activity of exudates from *T. pseudonana*, the presence of similar material in Arctic microlayer samples and the results of the filtering and heating experiments, we suggest that biogenic INPs present in phytoplankton exudates are a good candidate for the source of activity observed in the sampled microlayers. A substantial mass fraction of submicrometre sea-spray aerosol is



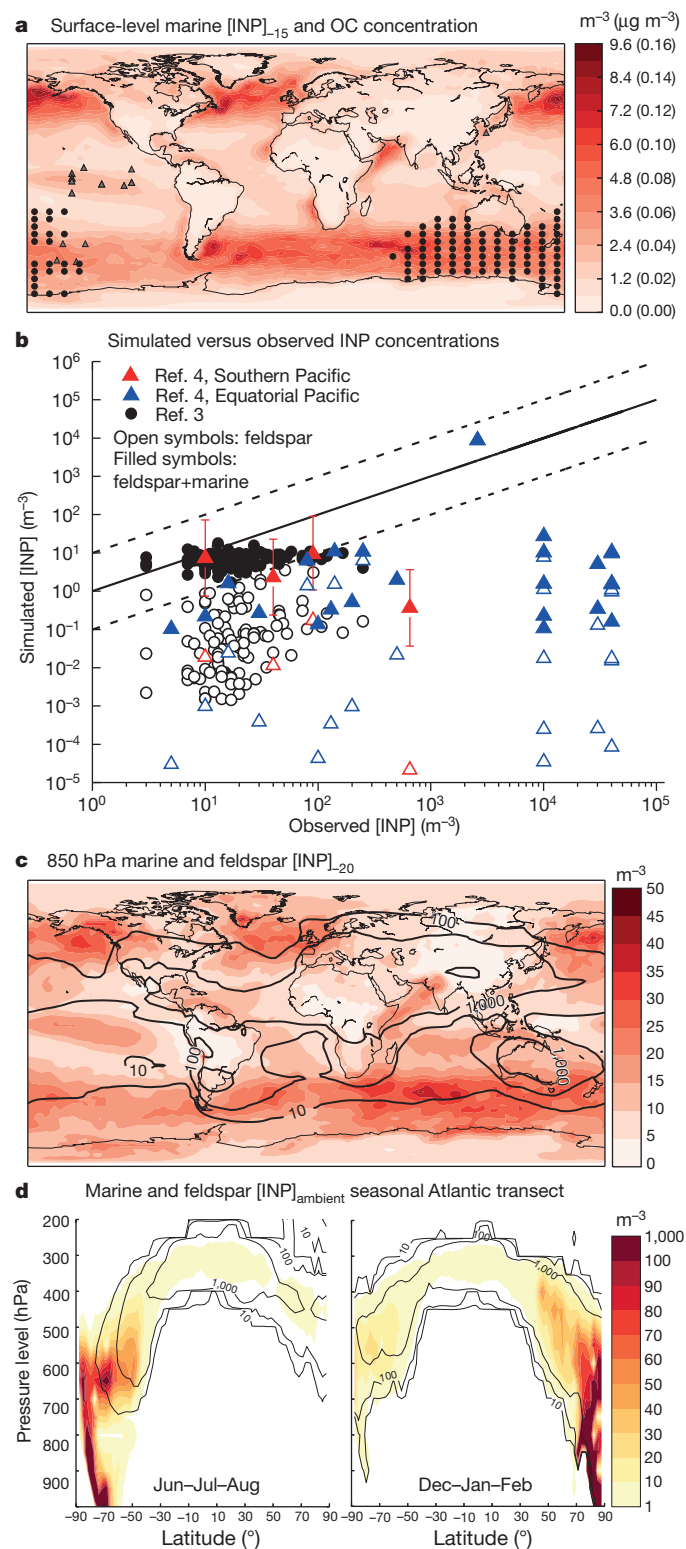
organic material<sup>12,16</sup>, which is often associated with phytoplankton exudates<sup>16,29</sup>. Our results indicate therefore that some fraction of sea-spray aerosol particles will be capable of nucleating ice.

To explore the possible contribution of marine biogenic INP sources to the global atmospheric INP distribution, we combined our experimental data with the modelled distribution of emitted primary organic material in sea-spray aerosol<sup>10</sup> (Fig. 4a; see Methods for details). To calculate atmospheric INP concentrations, we assume that the organic component of sea-spray aerosol simulated by the model has

a temperature-dependent INP concentration (per mass of organic carbon) equal to that measured in the Arctic and Atlantic samples (Fig. 2b). In other words, the number of INPs is directly related to the mass of organic carbon computed by the model. At our current level of understanding, this can be considered as an estimate for the number of INPs present in the organic component of sea-spray aerosol. The predicted surface-level marine organic INP concentrations ( $[INP]_{-15}$ ; the concentration of INPs if an air parcel were cooled to  $-15^{\circ}\text{C}$  at water saturation) are also shown in Fig. 4a. The largest concentrations of marine INPs occur over the Southern Oceans, the North Atlantic and the North Pacific in regions where biological activity in surface ocean waters and wind speeds are greatest. Comparison of the simulated annual mean marine INP concentrations at sea level agree within  $\pm 1$  order of magnitude with the Bigg<sup>3</sup> atmospheric INP measurements in the Southern Ocean and around the coast of Australia (Fig. 4b), while mineral dust from deserts (the major terrestrial INP source) only accounts for a small fraction of the observed INPs in this region. We also find good agreement with measurements made previously<sup>4</sup> in the southern Pacific. We note that some of the INP measurements made in the equatorial Pacific<sup>4</sup> tend to be under-predicted by our model, implying either another source of INPs or a stronger marine INP source than predicted here, possibly related to short-term variability in ocean biota.

We also use the model to assess the transport of marine INPs to altitudes relevant for mixed-phase clouds. Figure 4c shows the concentration of INPs active at  $-20^{\circ}\text{C}$  from marine sources at 850 hPa in comparison with the contribution from desert dusts based on K-feldspar distributions<sup>30</sup>. This suggests that marine biogenic sources of INPs are competitive with, or more important than, desert sources in large parts of the Southern Ocean, the North Atlantic and the North Pacific. To assess whether marine organic INPs exist in regions of the atmosphere that are sufficiently cold for them to activate to ice, we plot 3-monthly averaged seasonal distributions of INPs active at the ambient temperature ( $[INP]_{\text{ambient}}$ ) along a transect through the Atlantic (at  $30^{\circ}\text{W}$ ; Fig. 4d). This plot suggests that marine organic sea spray may contribute markedly to cloud glaciation at high and mid latitudes during the wintertime in the respective hemispheres. It also shows the strong seasonal differences caused by temperature and source strength to the relative  $[INP]_{\text{ambient}}$  contribution from marine and dust sources.

We show that surface active organic material from the microlayer that is similar to that found in sea-spray aerosol nucleates ice, and we tentatively identify the ice active material as being connected to diatom exudates. Our findings also suggest that marine organic material may be an important global source of atmospheric INPs, particularly in areas remote from terrestrial sources such as the Southern Ocean. This work highlights the need for more field measurements of remote atmospheric INP concentration and investigation of its relationship with ocean surface water characteristics, including the local phytoplankton community, organic carbon concentrations and



**Figure 4 | Global distribution of atmospheric marine biogenic INPs.**

**a**, Modelled distribution of INP concentration active at  $-15^{\circ}\text{C}$  ( $\text{m}^{-3}$ ) and surface-level marine aerosol organic mass concentration ( $\mu\text{g m}^{-3}$ ); the locations of Bigg<sup>3</sup> (circles) and Rosinski<sup>4</sup> (triangles) data are shown. **b**, Comparison of model-simulated INP concentration versus the Bigg<sup>3</sup> and Rosinski<sup>4</sup> measured concentrations for the same location at the activation temperature of the measurements; open symbols are for K-feldspar only and filled symbols are for mineral dust plus marine organic. Example error bars shown on red points are based on spread in INPs per gram TOC (see Fig. 2b). **c**, Modelled distribution of marine biogenic INP concentrations active at  $-20^{\circ}\text{C}$  at 850 hPa (corresponding to the altitude of high-latitude mixed-phase clouds). Black contours indicate the INPs from desert dust based on K-feldspar emissions<sup>30</sup>. **d**, Seasonal altitude profiles (expressed as pressure) showing  $[INP]_{\text{ambient}}$  (INP concentration active at local temperature conditions) from marine sources (colour scale) and K-feldspar (black contours), for a transect from the South to North poles through the Atlantic ( $30^{\circ}\text{W}$ ).

chemical characteristics, as well as the organic loading and nature of sea-spray aerosol.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 9 October 2014; accepted 17 July 2015.**

- Hoose, C. & Möhler, O. Heterogeneous ice nucleation on atmospheric aerosols: a review of results from laboratory experiments. *Atmos. Chem. Phys.* **12**, 9817–9854 (2012).
- Murray, B. J., O'Sullivan, D., Atkinson, J. D. & Webb, M. E. Ice nucleation by particles immersed in supercooled cloud droplets. *Chem. Soc. Rev.* **41**, 6519–6554 (2012).
- Bigg, E. K. Ice nucleus concentrations in remote areas. *J. Atmos. Sci.* **30**, 1153–1157 (1973).
- Rosinski, J., Haagenson, P. L., Nagamoto, C. T. & Parungo, F. Nature of ice-forming nuclei in marine air masses. *J. Aerosol Sci.* **18**, 291–309 (1987).
- Bigg, E. K. Ice forming nuclei in the high Arctic. *Tellus* **B48**, 223–233 (1996).
- Knopf, D. A., Alpert, P. A., Wang, B. & Aller, J. Y. Stimulation of ice nucleation by marine diatoms. *Nature Geosci.* **4**, 88–90 (2011).
- Fall, R. & Schnell, R. C. Association of an ice-nucleating pseudomonad with cultures of the marine dinoflagellate, *Heterocapsa niei*. *J. Mar. Res.* **43**, 257–265 (1985).
- Prather, K. A. *et al.* Bringing the ocean into the laboratory to probe the chemical complexity of sea spray aerosol. *Proc. Natl Acad. Sci. USA* **110**, 7550–7555 (2013).
- Yun, Y. & Penner, J. E. An evaluation of the potential radiative forcing and climatic impact of marine organic aerosols as heterogeneous ice nuclei. *Geophys. Res. Lett.* **40**, 4121–4126 (2013).
- Burrows, S. M., Hoose, C., Pöschl, U. & Lawrence, M. G. Ice nuclei in marine air: biogenic particles or dust? *Atmos. Chem. Phys.* **13**, 245–267 (2013).
- Wang, X. *et al.* Microbial control of sea spray aerosol composition: a tale of two blooms. *ACS Cent. Sci.* **1**, 124–131 (2015).
- Gantt, B. & Meskhidze, N. The physical and chemical characteristics of marine primary organic aerosol: a review. *Atmos. Chem. Phys.* **13**, 3979–3996 (2013).
- Aller, J. Y., Kuznetsova, M. R., Jahns, C. J. & Kemp, P. F. The sea surface microlayer as a source of viral and bacterial enrichment in marine aerosols. *J. Aerosol Sci.* **36**, 801–812 (2005).
- Orellana, M. V. *et al.* Marine microgels as a source of cloud condensation nuclei in the high Arctic. *Proc. Natl Acad. Sci. USA* **108**, 13612–13617 (2011).
- Schmitt-Kopplin, P. *et al.* Dissolved organic matter in sea spray: a transfer study from marine surface water to aerosols. *Biogeosciences* **9**, 1571–1582 (2012).
- Russell, L. M., Hawkins, L. N., Frossard, A. A., Quinn, P. K. & Bates, T. S. Carbohydrate-like composition of submicron atmospheric particles and their production from ocean bubble bursting. *Proc. Natl Acad. Sci. USA* **107**, 6652–6657 (2010).
- Leck, C. & Bigg, E. K. Biogenic particles in the surface microlayer and overlaying atmosphere in the central Arctic Ocean during summer. *Tellus* **B57**, 305–316 (2005).
- Cunliffe, M. *et al.* Sea surface microlayers: a unified physicochemical and biological perspective of the air–ocean interface. *Prog. Oceanogr.* **109**, 104–116 (2013).
- Quinn, P. K. *et al.* Contribution of sea surface carbon pool to organic matter enrichment in sea spray aerosol. *Nature Geosci.* **7**, 228–232 (2014).
- Cziczo, D. J. *et al.* Clarifying the dominant sources and mechanisms of cirrus cloud formation. *Science* **340**, 1320–1324 (2013).
- Yakobi-Hancock, J. D., Ladino, L. A. & Abbatt, J. P. D. Feldspar minerals as efficient deposition ice nuclei. *Atmos. Chem. Phys.* **13**, 11175–11185 (2013).
- Kanji, Z. A., Welti, A., Chou, C., Stetzer, O. & Lohmann, U. Laboratory studies of immersion and deposition mode ice nucleation of ozone aged mineral dust particles. *Atmos. Chem. Phys.* **13**, 9097–9118 (2013).
- Pummer, B. G., Bauer, H., Bernardi, J., Bleicher, S. & Grothe, H. Suspendable macromolecules are responsible for ice nucleation activity of birch and conifer pollen. *Atmos. Chem. Phys.* **12**, 2541–2550 (2012).
- Augustin, S. *et al.* Immersion freezing of birch pollen washing water. *Atmos. Chem. Phys.* **13**, 10989–11003 (2013).
- O'Sullivan, D. *et al.* The relevance of nanoscale biological fragments for ice nucleation in clouds. *Sci. Rep.* **5**, 8082 (2015).
- Fröhlich-Nowoisky, J. *et al.* Ice nucleation activity in the widespread soil fungus *Mortierella alpina*. *Biogeosciences* **12**, 1057–1071 (2015).
- Armbrust, E. V. The life of diatoms in the world's oceans. *Nature* **459**, 185–192 (2009).
- Alvain, S., Moulin, C., Dandonneau, Y. & Loisel, H. Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: a satellite view. *Glob. Biogeochem. Cycles* **22**, GB3001 (2008).
- Fuentes, E., Coe, H., Green, D. & McFiggans, G. On the impacts of phytoplankton-derived organic matter on the properties of the primary marine aerosol—part 2: composition, hygroscopicity and cloud condensation activity. *Atmos. Chem. Phys.* **11**, 2585–2602 (2011).
- Atkinson, J. D. *et al.* The importance of feldspar for ice nucleation by mineral dust in mixed-phase clouds. *Nature* **498**, 355–358 (2013).
- Koop, T., Luo, B., Tsias, A. & Peter, T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature* **406**, 611–614 (2000).
- Alpert, P. A., Aller, J. Y. & Knopf, D. A. Ice nucleation from aqueous NaCl droplets with and without marine diatoms. *Atmos. Chem. Phys.* **11**, 5539–5555 (2011).

**Acknowledgements** T.W.W., B.J.M. and T.F.W. acknowledge the assistance provided by the crew and other scientists onboard the R/V *Knorr* and the RRS *James Clark Ross*, the British Antarctic Survey, K. Baustian, J. McQuaid, A. Windross, J. Knulst, J. F. Wilson, A. M. Booth, R. Chance, L. J. Carpenter, S. Peppe, D. O'Sullivan, N. Umo, I. Cotton, H. Pearce, H. Price and M. J. Callaghan. The STXM/NEXAFS analysis was performed at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract no. DE-AC02-05CH11231 (user award to D.A.K./J.Y.A. ALS-05955). STXM analyses were facilitated by A. L. D. Kilcoyne and M. K. Gilles. L.A.L. acknowledges assistance from R. Leitch, E. Mungall, R. Christensen and J. Li, and the Pacific region Department of Fisheries and Oceans staff. The Marine Boundary Layer sampling site in Ucluelet is jointly supported and maintained by Environment Canada, the British Columbia Ministry of the Environment and Metro Vancouver. We acknowledge funding from the Natural Environment Research Council (NE/K004417/1, NE/I020059/1, NE/I013466/1, NE/I028696/1, NE/I019057/1, NE/H009485/1), the European Research Council (FP7, 240449 ICE, BACCHUS 603445), the UK Aerosol Society, National Science Foundation (AGS-1232203), German Research Foundation (WU585/6-1), the Climate Change and Atmospheric Research Program of the Natural Sciences and Engineering Research Council of Canada (for NETCARE), Fisheries and Oceans Canada, Environment Canada, NOAA's Climate Program Office (for WACS II), and the DOE Office of Science (BER) Earth System Modeling Program.

**Author Contributions** T.W.W. organized the ICE-ACCACIA campaign, designed experiments, collected and analysed samples during and after the campaign, managed collaborations and co-wrote this manuscript. L.A.L. designed experiments and analysed samples during the NETCARE campaign and co-wrote this manuscript. P.A.A. collected and analysed STXM/NEXAFS spectra and diatom exudate freezing data and contributed to manuscript writing. M.N.B. collected flow cytometry data for SML samples during ACCACIA. I.M.B. sought funding for ACCACIA and helped design the microlayer sampling procedure. S.M.B., J.V.T., J.B. and K.S.C. performed and analysed the model simulations. C.J. performed heating tests on ICE-ACCACIA samples. W.P.K. assisted with collection of material during the WACS II cruise, provided exudate material for experiments, and participated in STXM/NEXAFS data collection. G.M., J.J.N. and S.R. conducted total organic carbon measurements on Arctic samples. L.A.M., O.W. and E.P. collected the Ucluelet, Line P and open ocean samples, and conducted the NETCARE biogeochemical analysis. C.L.S. helped organize the measurements at the Ucluelet site and facilitated the use of the sampling site. T.F.W. collected and analysed samples during and after the ICE-ACCACIA campaign and assisted with design of experiments. J.D.Y.-H., J.A.H., R.H.M., M.S. and J.P.S.W. collected the Ucluelet samples and helped with the NETCARE experiments. A.K.B. and J.P.D.A. oversaw and organized the NETCARE field campaign and provided financial support for it. D.A.K. and J.Y.A. initiated and designed the STXM/NEXAFS and diatom exudate freezing experiments, contributed to the writing of this manuscript and provided financial support for WACS II cruise participation, exudate freezing experiments, and STXM/NEXAFS analyses. B.J.M. established the collaborations necessary for this paper, helped to write the paper and oversaw the ICE-ACCACIA campaign and sought funding for it.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.W.W. ([theo.w.wilson@gmail.com](mailto:theo.w.wilson@gmail.com)) or L.A.L. ([luis.ladinomoren@utoronto.ca](mailto:luis.ladinomoren@utoronto.ca)).

## METHODS

**SML and SSW sampling.** SML sampling took place in the Arctic during the Aerosol-Cloud Coupling and Climate Interactions in the Arctic (ACCACIA) cruise, in the Atlantic as part of the Western Atlantic Climate Study II (WACS II) cruise, as well as in the Northeast Pacific and off the southern coast of British Columbia, Canada, as part of the Network on Climate and Aerosols: Addressing Key Uncertainties in Remote Canadian Environments (NETCARE) project (Extended Data Fig. 1). See Extended Data Table 1 for precise sampling locations.

During the ACCACIA campaign in the Arctic, microlayer sampling was conducted from the RRS *James Clark Ross* in both open waters and within leads in the marginal ice zone (Extended Data Table 1 and Extended Data Fig. 1a). Microlayer samples were collected into borosilicate glass bottles from a hydrophilic Teflon film on a rotating drum fitted to the 'Interface II' remote-controlled sampling catamaran<sup>33</sup>. SSW was sampled with Niskin bottles on a CTD (conductivity, temperature, depth) rig at the same locations, generally at a depth of ~2 m. To avoid potential contamination from the ship, the microlayer sampler was navigated to a distance of 75–200 m upwind of the stationary ship before sampling commenced. Owing to rougher conditions during the WACS II campaign in the Atlantic, the Interface II was tethered to the CTD arm of the R/V *Knorr* during microlayer sampling. SSW was collected either using the direct uncontaminated ship input at 5-m water depth or using a sampling container lowered over the side of the ship (for details see Extended Data Table 1). During both campaigns, before and after microlayer sampling, sea water from the ship's uncontaminated supply was flushed through Interface II's sampling system to clear any previously collected microlayer. Samples of the 'flushing' water were collected and analysed and compared to the SSW using droplet freezing assays to check that INPs from previous sampling had been removed (Extended Data Fig. 8). Microlayer samples for water activity and organic carbon analysis were frozen immediately after collection at  $-80^{\circ}\text{C}$ .

During the NETCARE campaign, SML samples and corresponding SSW samples from water depths of 0.5–1.0 m were collected in the Northeast Pacific Ocean at three different locations (Extended Data Table 1 and Extended Data Fig. 1c). The samples were collected using a glass plate<sup>34</sup> with the exception of British Columbia (BC) coast SML3, which was collected using an autoclaved stainless steel screen<sup>18</sup>. All samples were stored in high-density polyethylene bottles. North Pacific samples (Pacific SML1, 2 and 3) were kept frozen at  $-20^{\circ}\text{C}$  after collection, and before the experiment they were thawed and stored at  $4^{\circ}\text{C}$  in the dark. All other samples were stored at  $4^{\circ}\text{C}$  in the dark for no more than 10 days before analysis.

**Effect of different sampling techniques on INP abundance in Pacific samples.** BC coast samples SML3 and SML4 were sampled at the same location 1 h apart but using metal mesh and glass plate techniques, respectively. There is a significant difference in the onset humidity of the two samples, with the onset of ice formation for BC coast SML3 occurring 13% RH<sub>ice</sub> lower than for BC coast SML4. With the time lag between collection of the two samples and because the methods sample different thicknesses of the SML (the metal plate collects layers 2–4 times thicker than those layers collected with glass plates<sup>18</sup>), it is not clear if the difference in their onset RH<sub>ice</sub> was due to the different sampling methods.

**Ice nucleation experiments during ACCACIA, WACS II and NETCARE.** Arctic and Atlantic SML and SSW samples were analysed for the presence of INPs using the previously described<sup>30,35</sup> microlitre nucleation by immersed particle instrument ( $\mu\text{L-NIPI}$ ). Briefly, droplets with a volume of  $1.0 \pm 0.1 \mu\text{L}$  were pipetted onto a hydrophobic microscope coverslip (Extended Data Fig. 9a) and cooled at a rate of  $1 \text{ K min}^{-1}$  using a Grant-Asymptote EF600 cold stage (Extended Data Fig. 9b) until all droplets were frozen. The temperature at which individual droplets froze was recorded, with an uncertainty of  $\pm 0.4^{\circ}\text{C}$ . Experiments were also performed using diluted microlayer samples (Arctic SML3, 6, 16 plus Arctic SML5 filtered at  $0.2 \mu\text{m}$  and Arctic SML12 filtered at  $10 \mu\text{m}$ ), where 1 ml of microlayer was added to 9 ml (10% dilution) or 99 ml (1% dilution) of  $18.2 \text{ M}\Omega \text{ cm}$  distilled water (Milli-Q). The water activity ( $a_w$ ) of Arctic microlayer and SSW samples was measured at  $25^{\circ}\text{C}$  using an Aqualab Series 3 dew point activity meter.

Samples from the North Pacific and BC coast were analysed using the University of Toronto continuous flow diffusion chamber (UT-CFDC). Owing to the high total submicrometre particle concentration of both the microlayer and SSW samples ( $10^6 \text{ cm}^{-3}$  after drying, measured with a TSI 3782 condensation particle counter) they were diluted by a factor of approximately 20 with  $18.2 \text{ M}\Omega \text{ cm}$  water before atomization using a TSI 3076 atomizer. Water was removed by passing the sample flow through three diffusion dryers, and the particle concentration was further lowered by dilution (Extended Data Fig. 9c). Two-hundred-nanometre mobility diameter particles were selected using a differential mobility analyser (TSI 3081), with size-selected particle concentrations of approximately  $100 \text{ cm}^{-3}$ . The size-selected particles were exposed to ice super-saturated conditions at  $-40^{\circ}\text{C}$  in the UT-CFDC to determine their ice nucleation

onset humidities<sup>36</sup>. Particle counts from both channels ( $>0.5 \mu\text{m}$  and  $>5 \mu\text{m}$ ) of the optical particle counter (Climet CI-20) were used to distinguish between interstitial aerosol particles and ice particles. The standard solutions consisted of 8–10 mg of NaCl (Sigma Aldrich, S2830) and commercial sea salt (Sigma Aldrich, S9883) dissolved in 50 ml of  $18.2 \text{ M}\Omega \text{ cm}$  water. Control experiments with filtered air were conducted in the field.

**Ice nucleation experiments with diatom exudates.** For diatom exudate freezing experiments, axenic unialgal cultures of *T. pseudonana* were grown in flasks at  $16\text{--}18^{\circ}\text{C}$  with a 14 h light:10 h dark cycle in  $0.1 \mu\text{m}$ -filtered and autoclaved sea water with f/2 nutrient addition<sup>37</sup>. Sea water was collected at a depth of about 0.5 m about 100 km off the coast of Long Island, New York<sup>38</sup>. After 1 week of growth, when concentrations reached  $\sim 10^6 \text{ cells ml}^{-1}$ , cultures were filtered through a  $0.1 \mu\text{m}$  pore size filter to remove the cells, yielding a suspension of diatom exudates.

Droplets of filtered diatom exudate were analysed using the water-activity-controlled immersion freezing experiment (WACIFE) instrument. Individual droplets were deposited in a grid pattern onto a hydrophobic glass plate. Additional droplets were generated from filtered and autoclaved natural sea water with and without f/2 nutrient addition<sup>37</sup> before diatom growth. Droplet  $a_w$  was established by allowing the temperature-controlled droplets to come to equilibrium in a humidity-controlled aerosol conditioning cell<sup>6,39</sup>. Droplets were then sealed from ambient air, setting the droplets'  $a_w$  equal to the applied RH. Ice nucleation was observed at a cooling rate of  $10 \text{ K min}^{-1}$  using a cryo-cooling stage coupled to an optical microscope (see schematic of process in Extended Data Fig. 9d). Droplet sizes ranged from 60 to  $129 \mu\text{m}$  circle equivalent diameters ( $82 \mu\text{m}$  average). Individual droplet volumes were calculated from the spherical equivalent diameter derived from the digitally measured particle diameters corrected for the non-sphericity of the deposited particles and for different applied  $a_w$ . The total number of droplets at each investigated  $a_w$  for sea water, sea water plus f/2 droplets, and sea water plus f/2 plus diatom exudates were 115, 143 and 131, respectively. Homogeneous ice nucleation was observed for droplets generated from the sea water with and without f/2. The median freezing temperatures shown in Fig. 3d include 10th and 90th percentiles. Corresponding mean melting temperatures are shown with an uncertainty of 1 standard deviation. The uncertainty in  $a_w$  is  $\pm 0.01$ . The ice melting curve (dashed line) and the volume-corrected homogeneous freezing curve with an uncertainty in  $a_w$  of  $\pm 0.01$  shown as the solid black line and grey shaded area, respectively, are parameterized as described previously<sup>40</sup>.

**Freezing depression correction for immersion mode experiments.** Data from the Arctic, Atlantic and diatom exudate immersion mode experiments were adjusted to account for the freezing depression caused by dissolved salts in sea water. First, a freezing curve as a function of  $a_w$  was constructed through the median freezing points following the  $a_w$  criterion where median immersion freezing temperatures can be described by a horizontal shift in the ice melting curve<sup>40</sup>. Then the difference between expected median freezing temperatures for pure water (that is, at  $a_w = 1.0$ ) and at the experimentally applied  $a_w$  were used as temperature offsets.

**Calculation of INPs per gram of organic carbon and associated uncertainty.** In Fig. 2b we show the cumulative number of INPs per gram of TOC as a function of temperature for the Arctic and Atlantic microlayer samples. This calculation uses the time-independent singular description of ice nucleation<sup>41</sup> that assumes the time dependence of freezing is of secondary importance to the distribution of ice nucleating particle types. In this case 'INPs per gram of TOC' is the same as ' $n_m$ ', described in detail elsewhere<sup>2</sup>. It should also be noted that this is the same model that was used to calculate  $n_s$  (ref. 1) for the BC coast and North Pacific samples, but this describes the number of ice active sites per surface area rather than mass.

For all data points shown in Fig. 2b, error bars are based on the propagated uncertainties associated with volume measurements and organic carbon concentration measurements (Extended Data Fig. 7). The error bars for experiments in which microlayer samples were diluted with Milli-Q water also include uncertainty relating to the subtraction of background heterogeneous nucleation events.

Freezing in  $\mu\text{L-NIPI}$  experiments using Milli-Q water droplets free from any added nucleants occurs at higher temperatures than predicted for homogeneous ice nucleation<sup>42</sup>. Using the results of 22 separate freezing experiments (727 droplets in total) the cumulative number of INPs per volume of Milli-Q was calculated. A line of best fit from this data as a function of temperature was used to estimate the number of background INPs present in our diluted microlayer samples. This value was subtracted from the diluted microlayer cumulative INP spectra and uncertainties relating to the variation in background INP concentrations were calculated based on the 68% confidence interval associated with the line of best fit.

**STXM/NEXAFS analysis of Arctic microlayer samples.** STXM/NEXAFS analysis was used to explore qualitatively the carbon functionality of organic material found in two Arctic SML samples (Arctic SML5 and 19). Analyses were performed at the Advanced Light Source on beamline 5.3.2.2, Lawrence Berkeley National



Laboratory<sup>43</sup>. Overviews of the application of this technique to atmospheric particles and technical details on STXM methodology have been published elsewhere<sup>44–50</sup>. The sample for STXM/NEXAFS analysis was collected by bringing the flat face of clean silicon nitride windows into contact with the pre-collected SML water surface and then lifting them off. Material adhering to the windows was allowed to air dry before examination. The transmission of soft single-energy X-ray photons across the raster-scanned sample was measured to obtain an image<sup>43</sup> exploiting the carbon K-edge spectra to identify carbon functionality and the overall contribution of inorganic components. Using X-ray energies 278–320 eV, X-ray absorption of the ground state electron (1s orbital) of the carbon atom was probed to identify carbon–carbon double bonding (C=C), carbonyl (C=O), hydroxyl (C–OH) and carboxyl (COOH) functional groups. Figure 3a shows an X-ray image of particulate material found in the Arctic SML5 sample taken at 320 eV, which includes organic and inorganic material and a frustule fragment. The spectrum for SML5 is characterized by (1) a dominant carbonyl peak (288.2 eV), (2) a secondary carbon double-bonding peak (285.4 eV), and (3) a gradual rise in absorption in the energy range where the hydroxyl functional group absorbs (287.3 eV). Similar spectra for diatom cell wall material have previously been observed<sup>51</sup>. The spectrum observed from SML19 is similar to that of the spectrum obtained from diatom exudates and is characterized by (1) a dominant carboxyl peak (288.6 eV) and (2) a gradual rise in absorption in the energy range where the hydroxyl functional group absorbs (287.3 eV). Similar spectral features have been observed previously<sup>52</sup> for field-collected marine particles, which the authors attributed to the presence of polysaccharides, and also<sup>53</sup> for particles aerosolized from a laboratory breaking wave. These features were also observed in carbon absorption spectra<sup>54</sup> for 22 amino acids, all of which are present in *T. pseudonana*<sup>55</sup>, supporting the notion that diatom exudates were present in this microlayer sample. The presence of spectral features similar to diatom cell wall fragments and diatom exudate material suggests the presence of diatoms in both SML5 and SML19. It is important to note that marine SMLs<sup>13,18,56–58</sup>, SSWs<sup>59</sup> and biofilms<sup>60</sup> are comprised of a complex mixture of inorganic particles, particulate organic matter in the form of microorganisms, biogenic debris, polysaccharide enriched microgels, lipids, proteins and amino acids. Therefore, the X-ray spectra shown here will not be identical for different SML samples. Instead, key spectral features including peak locations and dominance, represent typical biogenically derived materials in the marine environment.

**Microlayer filtering and heating tests.** A selection of fresh Arctic SML samples were passed through filters with a range of pore sizes (0.02 µm Whatman Anodisc, 0.1 µm Whatman Anodisc, 0.2 µm Sartorius Minisart, 0.45 µm Sartorius Minisart, 2.0 µm Millipore TTP) and then tested for changes in immersion mode ice nucleating activity (Extended Data Fig. 2a). Selected NETCARE samples were filtered through 0.2 µm polyethersulfone membranes (IC Acrodisc) and then retested under cirrus conditions for ice nucleation activity (Extended Data Fig. 2b).

Selected Arctic and Atlantic samples were tested for immersion freezing activity after having been heated in a temperature-controlled bath for 10 min. This was repeated at 8 temperatures between 20 °C and 100 °C (Extended Data Fig. 2c). NETCARE samples, BC coast SML3 and Pacific SML1 were heated for 10 min at 100 °C and retested under cirrus conditions for ice nucleation activity (Extended Data Fig. 2b).

**Flow cytometry for ACCACIA SML samples.** Samples (2 ml) were transferred into a cryovial, and 50 µl of 50% glutaraldehyde was added to achieve a 0.5% solution. The preserved sample was stored in the refrigerator for 30 min before snap-freezing in liquid nitrogen for storage at –80 °C. Prior to analysis, samples were defrosted and the nucleic acid stain SYBR Green 1 dissolved in 300 mM potassium nitrate solution was added to achieve a 1% concentration of the stain (see refs 61 and 62 for more details). The samples were kept in the dark at room temperature for 1 h. Bacterial abundance shown in Extended Data Fig. 4 was quantified using a flow cytometer (Becton Dickinson FACScan) following methods outlined in the literature<sup>61,62</sup>.

**TOC, DOC and TEP characterization.** Transparent exopolymer particle (TEP; Extended Data Fig. 5b) concentrations were determined using spectrophotometric methods<sup>63</sup>. Samples (5–50 ml) of NETCARE SML and SSW were filtered immediately after sampling onto 0.2 µm polycarbonate membranes under low vacuum pressure (<100 mm Hg) and the membranes were stained with 0.5 ml Alcian blue solution (0.02 g Alcian blue in 100 ml of acetic acid solution, pH 2.5), and rinsed twice with 1 ml of deionized water. Membranes were subsequently extracted in 6 ml of 80% sulfuric acid (H<sub>2</sub>SO<sub>4</sub>) for 2 h to dissolve the dye, and the absorbance of extracts was measured in a 1 cm cuvette at 787 nm with standardization using xanthan gum equivalents (X<sub>eq</sub>) and conversion to µg C l<sup>–1</sup> using a factor of 0.63 based on the compilation of multiple studies using phytoplankton cultures<sup>64</sup>.

For analysis of DOC<sup>65</sup> in the NETCARE samples from the North Pacific (Extended Data Fig. 5a), subsamples of 20 ml were filtered through 0.2 µm polycarbonate membranes and the filtrates preserved for later analysis with 4 µl

H<sub>3</sub>PO<sub>4</sub> per ml sample. Concentrations of DOC were quantified using a Shimadzu TOC-V analyser.

Organic carbon analysis was also performed on the Arctic microlayer samples (Extended Data Fig. 7). After filtering through 0.2 µm pore size Whatman Anodisc filters, DOC was measured using a Shimadzu TOC-V using previously described methods<sup>66,67</sup>. Owing to high carbon concentration the Arctic microlayer samples for total organic carbon (TOC) analysis were diluted with one part sample to two parts deionized water. Total carbon analyser (TOC 5050A, Shimadzu) was used to measure total organic and inorganic carbon in each sample twice (with coefficient of variance between measurements <2%); the average of these measurements was taken.

Atlantic samples were also analysed for DOC; 40 ml of sample was filtered through a 25 mm GFF filter and stored in a glass septum bottle. The resulting supernatant was acidified to a pH of 1 with roughly 3 drops of pure HCl to react with any inorganic carbon, and inhibit potential microbial degradation before analysis in the DOC analyser (Shimadzu TOC-L, ±1% precision). Prior to DOC collection each septum bottle and filtration apparatus was acid washed, and GFF filters were pre-combusted for 5 h at 450 °C. Particulate organic carbon (POC) was measured in Atlantic samples by filtering SML through pre-combusted 25 mm GFF filters (0.7 µm pore size) before analysis (filters were frozen at –20 °C until processing).

**Global modelling study.** To assess the potential contribution of marine biogenic INP sources to the global atmospheric INP distribution, we combined our experimental data from immersion mode experiments using the Arctic and Atlantic SML samples (the fit shown in Fig. 2b) with a modelled distribution of emitted primary organic material in sea-spray aerosol<sup>10</sup>. This was compared to desert dust INP concentrations based on emissions of K-feldspar<sup>30</sup>.

Atmospheric marine POC distributions are taken from ref. 10. Briefly, organic material (OM) in sea spray is related to emitted NaCl by an empirical, observationally based enrichment factor ( $(OM_{aerosol}/NaCl_{aerosol})/(OM_{sea}/NaCl_{sea}) = 500$ ). Furthermore, the OM fraction in emitted sea spray ( $(OM_{aerosol}/(OM_{aerosol} + NaCl_{aerosol}))$ ) was not allowed to exceed 76%; which was the maximum observed submicrometre organic sea spray fraction according to ref. 68. Marine organic carbon is emitted in proportion to climatological POC as retrieved by the MODIS-Aqua satellite, transported as soluble  $r = 100$  nm particles in the atmosphere, and removed by size-dependent wet and dry deposition processes. The OM distribution (using  $OM = OC \times 1.9$ ) was shown previously<sup>10</sup> to be consistent with another global model study of atmospheric marine OM<sup>68</sup> and is within a factor of two of annual mean atmospheric measurements at Amsterdam Island (37°48' S, 77°34' E)<sup>69</sup> and Mace Head (53°20' N, 9°54' W)<sup>68</sup>.

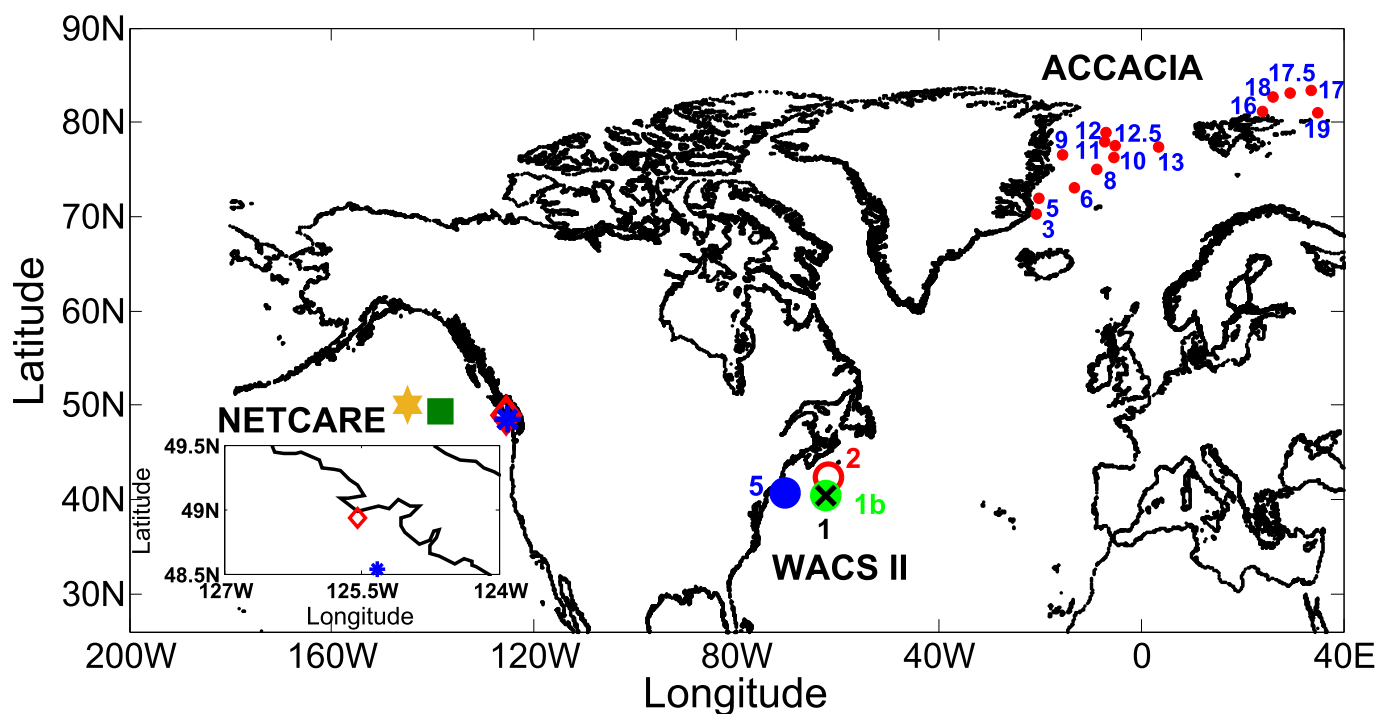
The distribution of dust INPs was based on a previous study<sup>30</sup>, which used the global aerosol processes model (GLOMAP). In this study we used GLOMAP-mode<sup>70</sup>, a two-moment aerosol size-resolving scheme, which calculates particle mass and number in seven variable-size log-normal modes. The model is forced by ECMWF 6-hourly global meteorological analyses and was run at a resolution of  $2.8 \times 2.8$ , with 31 pressure levels extending to ~10 hPa. Dust emissions are taken from the AEROCOM daily resolved dust inventory for 2000 (ref. 71). Emissions are separated into a feldspar and bulk component using a mineralogical inventory<sup>72</sup>. Dust is emitted into the insoluble (or primary) size distribution and is aged to the soluble distribution via condensation of SO<sub>2</sub> and secondary organics after which it is subject to wet scavenging processes. Evaluation of modelled dust mass concentrations against surface observations (from the University of Miami network) show a model bias of ~30% with the majority of observations within a factor of 2 of the modelled mass<sup>70</sup>.

Mineral dust INP concentrations (originating from K-feldspar) were calculated offline via the time independent model as discussed previously<sup>30</sup>. K-Feldspar (assumed to be 35% of Feldspar volume) surface area and particle number was calculated assuming external mixing within the soluble modes.

Figure 4d shows seasonal modelled concentrations of marine (colour scale) and K-feldspar (contours) INPs that are active at the ambient model temperature. We refer to this concentration as  $[INP]_{ambient}$  and it is a useful indicator of locations in which INP concentrations are sufficiently high and the temperatures sufficiently low to potentially influence clouds. Both marine and K-feldspar  $[INP]_{ambient}$  were calculated using averaged daily temperatures and averaged monthly marine organic and K-feldspar emissions during the indicated periods (left panel, December–January–February; right panel, June–July–August). In both cases the parameterizations used are valid over a limited temperature range (see Fig. 2b for marine INP, and ref. 30 for K-feldspar). We did not extrapolate for INP concentrations at temperatures above the upper limits of the parameterization, instead we assumed the aerosol had no ice-nucleating activity at higher temperatures. Owing to the INP numbers being cumulative as temperature decreases, the concentration of INPs at the lowest valid temperature of the parameterizations represent the lower limits for INP concentrations at temperatures colder than the valid range.

Below  $-38^{\circ}\text{C}$  we do not show  $[\text{IN}]_{\text{ambient}}$  since in this regime homogeneous nucleation will dominate.

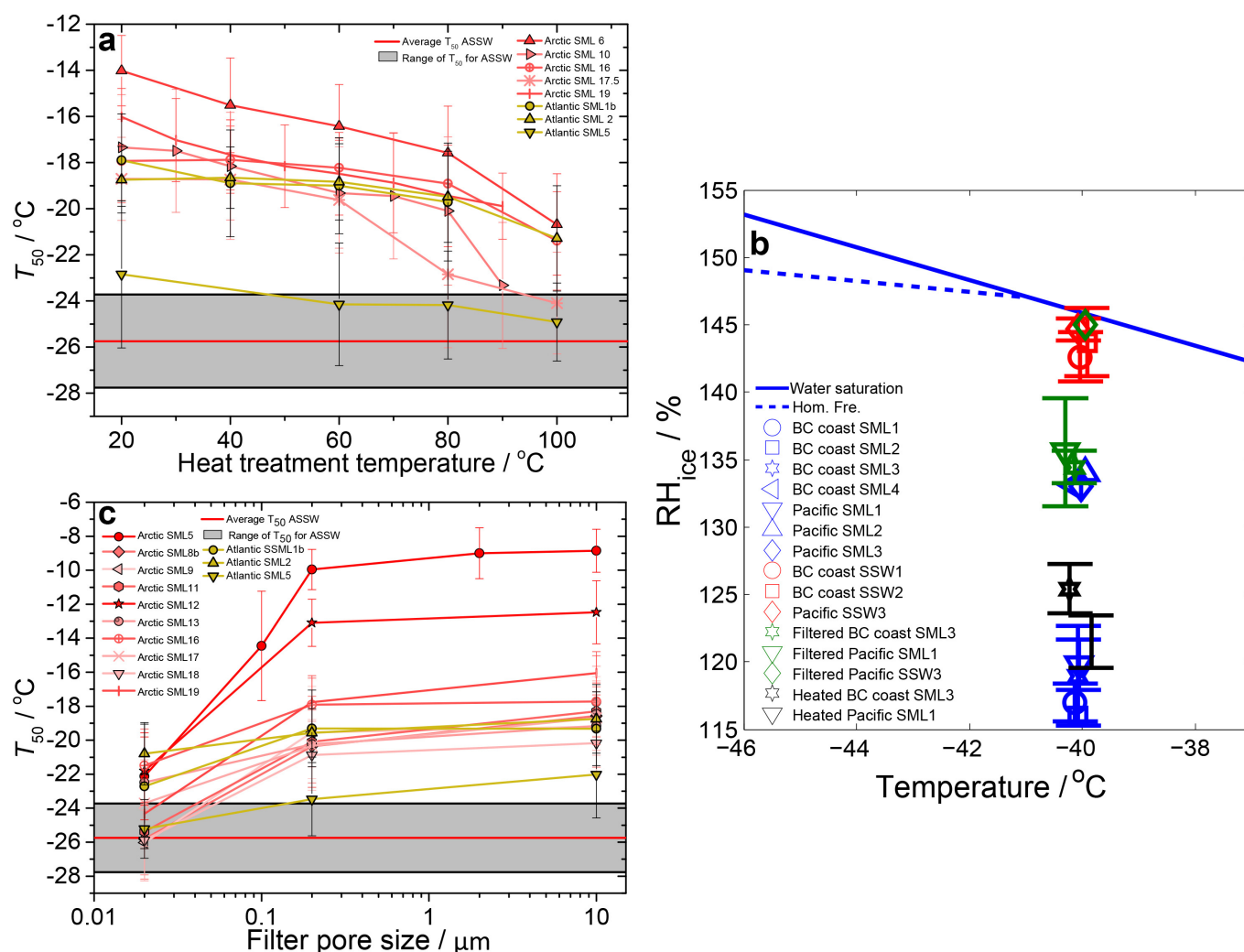
33. Knulst, J. C., Rosenberger, D., Thompson, B. & Paatero, J. Intensive sea surface microlayer investigations of open leads in the pack ice during Arctic Ocean 2001 expedition. *Langmuir* **19**, 10194–10199 (2003).
34. Wurl, O., Miller, L., Röttgers, R. & Vagle, S. The distribution and fate of surface-active substances in the sea-surface microlayer and water column. *Mar. Chem.* **115**, 1–9 (2009).
35. Whale, T. F. *et al.* A technique for quantifying heterogeneous ice nucleation in microlitre supercooled water droplets. *Atmos. Meas. Tech.* **8**, 2437–2447 (2015).
36. Kanji, Z. A. & Abbatt, J. P. D. The University of Toronto Continuous Flow Diffusion Chamber (UT-CFDC): a simple design for ice nucleation studies. *Aerosol Sci. Technol.* **43**, 730–738 (2009).
37. Guillard, R. R. L. & Ryther, J. H. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Grun. *Can. J. Microbiol.* **8**, 229–239 (1962).
38. Fisher, N. S. & Wente, M. The release of trace elements by dying marine phytoplankton. *Deep-Sea Res. I* **40**, 671–694 (1993).
39. Alpert, P. A., Aller, J. Y. & Knopf, D. A. Initiation of the ice phase by marine biogenic surfaces in supersaturated gas and supercooled aqueous phases. *Phys. Chem. Chem. Phys.* **13**, 19882–19894 (2011).
40. Koop, T. & Zobrist, B. Parameterizations for ice nucleation in biological and atmospheric systems. *Phys. Chem. Chem. Phys.* **11**, 10839–10850 (2009).
41. Vali, G. Quantitative evaluation of experimental results on the heterogeneous freezing nucleation of supercooled liquids. *J. Atmos. Sci.* **28**, 402–409 (1971).
42. Murray, B. J. *et al.* Kinetics of the homogeneous freezing of water. *Phys. Chem. Chem. Phys.* **12**, 10380–10387 (2010).
43. Kilcoyne, A. L. D. *et al.* Interferometer-controlled scanning transmission X-ray microscopes at the Advanced Light Source. *J. Synchrotron Radiat.* **10**, 125–136 (2003).
44. Ghorai, S., Laskin, A. & Tivanski, A. V. Spectroscopic evidence of keto-enol tautomerism in deliquesced malonic acid particles. *J. Phys. Chem. A* **115**, 4373–4380 (2011).
45. Ghorai, S. & Tivanski, A. V. Hygroscopic behavior of individual submicrometer particles studied by X-ray spectromicroscopy. *Anal. Chem.* **82**, 9289–9298 (2010).
46. Moffet, R. C., Henn, T., Laskin, A. & Gilles, M. K. Automated chemical analysis of internally mixed aerosol particles using X-ray spectromicroscopy at the carbon K-edge. *Anal. Chem.* **82**, 7906–7914 (2010).
47. Moffet, R. C. & Tivanski, A. V. & Gilles, M. K. in *Fundamentals and Applications of Aerosol Spectroscopy* (eds Signorell, R. & Reid, J. P.) 419–462 (Taylor & Francis Group, 2010).
48. Hopkins, R. J., Tivanski, A. V., Marten, B. D. & Gilles, M. K. Chemical bonding and structure of black carbon reference materials and individual carbonaceous atmospheric aerosols. *J. Aerosol Sci.* **38**, 573–591 (2007).
49. Knopf, D. A. *et al.* Microspectroscopic imaging and characterization of individually identified ice nucleating particles from a case field study. *J. Geophys. Res. Atmos.* **119**, 10365–10381 (2014).
50. Takahama, S., Gilardoni, S., Russell, L. M. & Kilcoyne, A. L. D. Classification of multiple types of organic carbon composition in atmospheric particles by scanning transmission X-ray microscopy analysis. *Atmos. Environ.* **41**, 9435–9451 (2007).
51. Abramson, L., Wirick, S., Lee, C., Jacobsen, C. & Brandes, J. A. The use of soft X-ray spectromicroscopy to investigate the distribution and composition of organic matter in a diatom frustule and a biomimetic analog. *Deep-Sea Res. II* **56**, 1369–1380 (2009).
52. Hawkins, L. N. & Russell, L. M. Polysaccharides, proteins, and phytoplankton fragments: four chemically distinct types of marine primary organic aerosol classified by single particle spectromicroscopy. *Adv. Meteorol.* **2010**, 612132 (2010).
53. Ault, A. P. *et al.* Size-dependent changes in sea spray aerosol composition and properties with different seawater conditions. *Environ. Sci. Technol.* **47**, 5603–5612 (2013).
54. Zubavichus, Y., Shaporenko, A., Grunze, M. & Zharnikov, M. Innershell absorption spectroscopy of amino acids at all relevant absorption edges. *J. Phys. Chem. A* **109**, 6998–7000 (2005).
55. Brown, M. R. The amino-acid and sugar composition of 16 species of microalgae used in mariculture. *J. Exp. Mar. Biol. Ecol.* **145**, 79–99 (1991).
56. Kuznetsova, M., Lee, C. & Aller, J. Characterization of the proteinaceous matter in marine aerosols. *Mar. Chem.* **96**, 359–377 (2005).
57. van Pinxteren, M., Müller, C., Iinuma, Y., Stolle, C. & Herrmann, H. Chemical characterization of dissolved organic compounds from coastal sea surface microlayers (Baltic Sea, Germany). *Environ. Sci. Technol.* **46**, 10455–10462 (2012).
58. Verdugo, P. Marine microgels. *Annu. Rev. Mar. Sci.* **4**, 375–400 (2012).
59. Brandes, J. A. *et al.* Examining marine particulate organic matter at sub-micron scales using scanning transmission X-ray microscopy and carbon X-ray absorption near edge structure spectroscopy. *Mar. Chem.* **92**, 107–121 (2004).
60. Lawrence, J. R. *et al.* Scanning transmission X-ray, laser scanning, and transmission electron microscopy mapping of the exopolymeric matrix of microbial biofilms. *Appl. Environ. Microbiol.* **69**, 5543–5554 (2003).
61. Marie, D., Partensky, F., Jacquet, S. & Vaulot, D. Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl. Environ. Microbiol.* **63**, 186–193 (1997).
62. Zubkov, M. V., Burkill, P. H. & Topping, J. N. Flow cytometric enumeration of DNA-stained oceanic planktonic protists. *J. Plankton Res.* **29**, 79–86 (2007).
63. Engel, A. in *Practical Guidelines for the Analysis of Seawater* (ed. Wurl, O.) 125–142 (CRC Press, 2009).
64. Engel, A. Distribution of transparent exopolymer particles (TEP) in the northeast Atlantic Ocean and their potential significance for aggregation processes. *Deep-Sea Res. I* **51**, 83–92 (2004).
65. Wurl, O. & Sin, T. in *Practical Guidelines for the Analysis of Seawater* (ed. Wurl, O.) 33–48 (CRC Press, 2009).
66. Pan, X. *et al.* Dissolved organic carbon and apparent oxygen utilization in the Atlantic Ocean. *Deep-Sea Res. I* **85**, 80–87 (2014).
67. MacGilchrist, G. A. *et al.* Effect of enhanced pCO<sub>2</sub> levels on the production of dissolved organic carbon and transparent exopolymer particles in short-term bioassay experiments. *Biogeosciences* **11**, 3695–3706 (2014).
68. Vignati, E. *et al.* Global scale emission and distribution of sea-spray aerosol: sea-salt and organic enrichment. *Atmos. Environ.* **44**, 670–677 (2010).
69. Sciare, J. *et al.* Long-term observations of carbonaceous aerosols in the Austral Ocean atmosphere: Evidence of a biogenic marine organic source. *J. Geophys. Res. Atmos.* **114**, D15302 (2009).
70. Mann, G. W. *et al.* Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model. *Geosci. Model Dev.* **3**, 519–551 (2010).
71. Huneus, N. *et al.* Global dust model intercomparison in AeroCom phase I. *Atmos. Chem. Phys.* **11**, 7781–7816 (2011).
72. Nickovic, S., Vukovic, A., Vujadinovic, M., Djurdjevic, V. & Pejanovic, G. Technical note: high-resolution mineralogical database of dust-productive soils for atmospheric dust modeling. *Atmos. Chem. Phys.* **12**, 845–855 (2012).



**Extended Data Figure 1 | Sampling locations.** SML and SSW samples were collected during the ACCACIA campaign (July–August 2013) at Arctic sampling stations at the locations marked with solid red circles. Also shown are sampling locations during the WACS II campaign (May–June 2014) in the North Atlantic Ocean. NETCARE samples were collected at locations in the

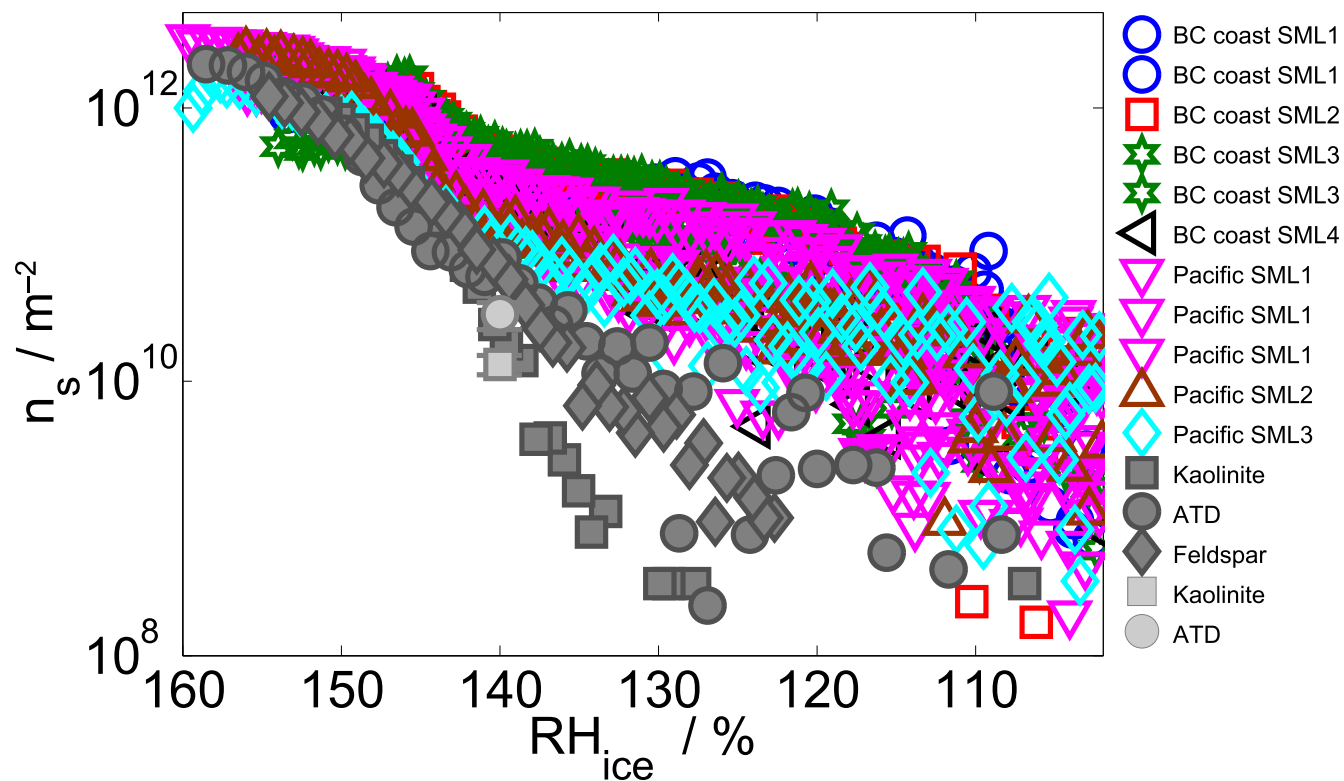
Northeast Pacific (yellow star and green square, CCGS *John P. Tully*, 14–19 June 2013). The red diamond and blue asterisk correspond to the sampling locations for the NETCARE British Columbia (BC) coastal samples (12–15 August 2013). The inset is a zoom of the BC coast sampling locations.





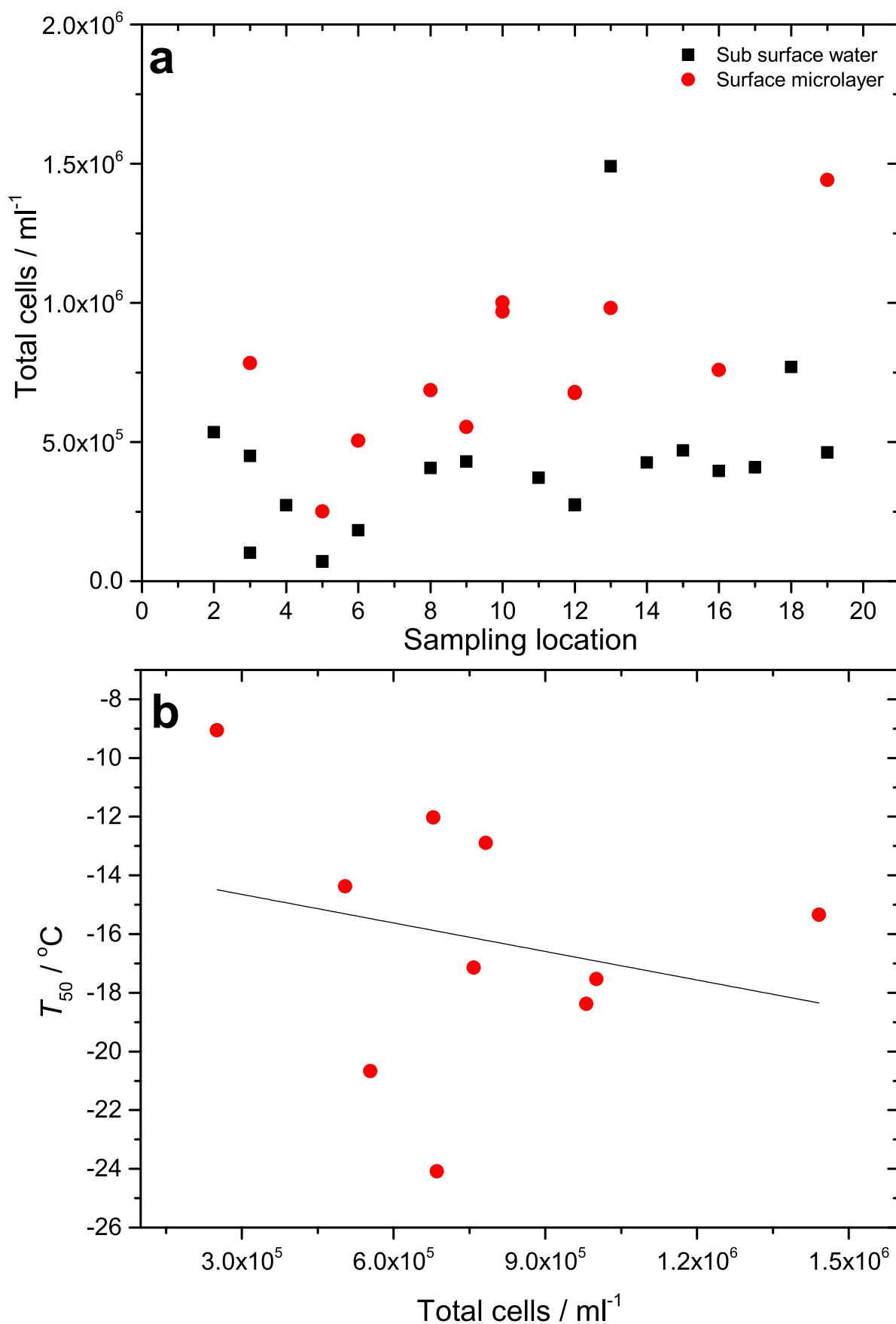
**Extended Data Figure 2 | Effects of heating and filtering on the ice nucleation activity of microlayer samples.** **a**, The effect of filtering through different pore-sized filters on the temperature at which 50% of droplets had frozen ( $T_{50}$ ) of Arctic and Atlantic SML samples tested using the  $\mu$ l-NIPI. Error bars represent  $\pm$  the standard deviation calculated from the freezing temperatures in each experiment, which consisted of between 30 and 53 individual events. Shaded grey area is the range of  $T_{50}$  found for fresh unfiltered SSW during the campaign. **b**, Comparison of the UT-CFDC onset  $RH_{ice}$  of unfiltered, filtered (0.2  $\mu$ m) and heated (to 100 °C for 10 min) North Pacific and

BC coast SML and SSW samples. The blue lines, and the red and dark blue symbols are as shown in Fig. 2d. The green symbols represent the filtered onsets, whereas the black symbols represent the heated results. Ice-nucleation-onset error bars represent one standard deviation based on three to four replicates. **c**, Results of heating tests using Arctic and Atlantic SML samples on  $T_{50}$  tested using the  $\mu$ l-NIPI. Error bars represent  $\pm$  the standard deviation calculated from the freezing temperatures in each experiment, which consisted of between 28 and 46 individual events. Shaded grey area is the range of  $T_{50}$  found for fresh untreated Arctic SSW.



**Extended Data Figure 3 | Ice surface site densities for the Pacific microlayer samples.** Comparison of the ice surface densities ( $n_s$ ) calculated from UT-CFDC data for the NETCARE SML samples with literature data. The  $n_s$  values were obtained at  $-40^\circ\text{C}$  assuming that the particles were spherical. The

SML  $n_s$  values are indicated by the coloured symbols, whereas the mineral dust  $n_s$  values are indicated by the grey symbols. The dark grey and light grey symbols are from refs 21 and 22, respectively.

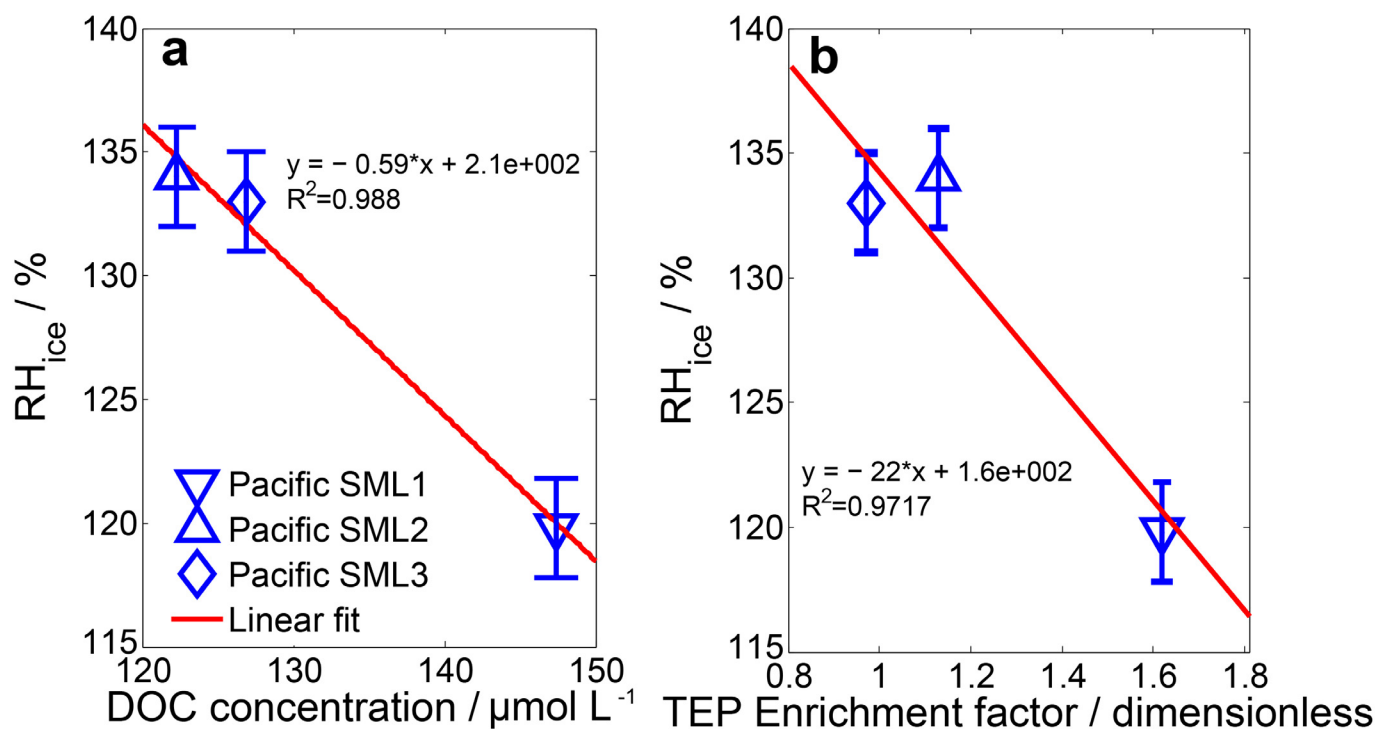


**Extended Data Figure 4 | Bacterial cell counts for Arctic samples.**

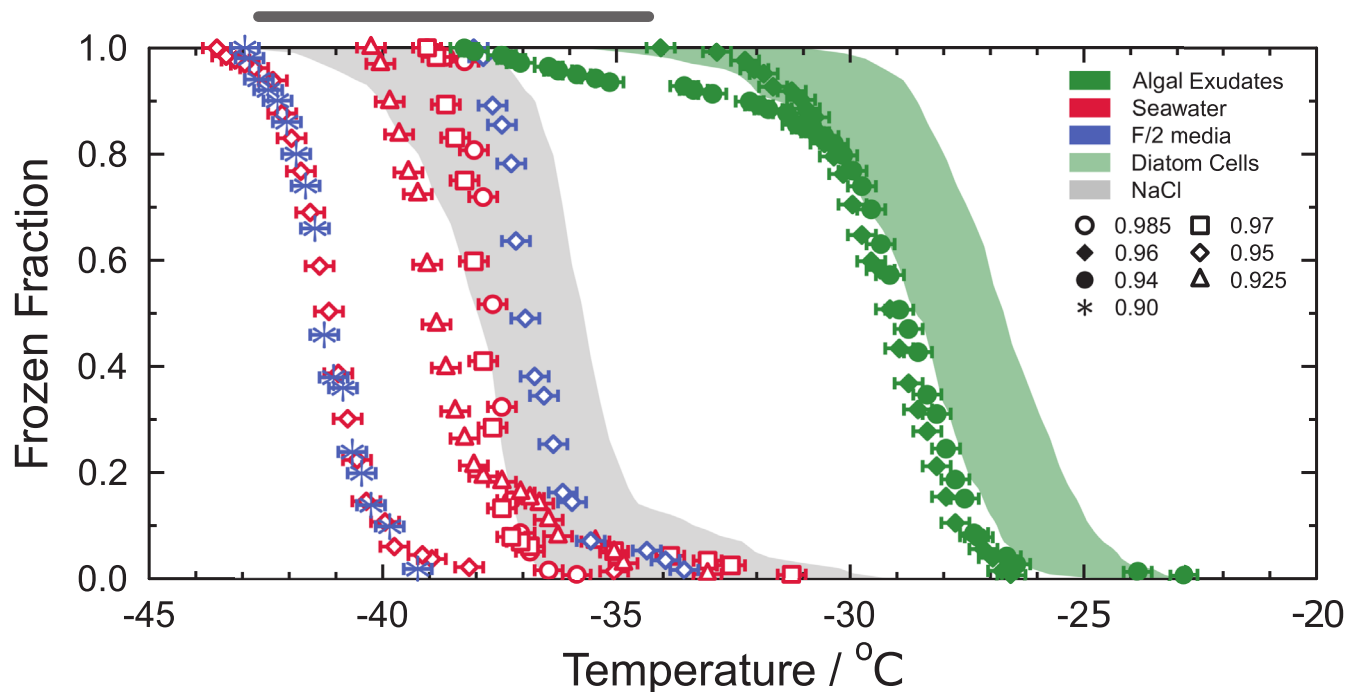
**a**, Bacterial cell counts from flow cytometry performed on Arctic SSW (black squares), fresh Arctic SML (red circles). **b**, The SML sample cell counts plotted

against  $T_{50}$  (temperature at which 50% of droplets frozen) and line of best fit,  $R^2 = 0.29$ .



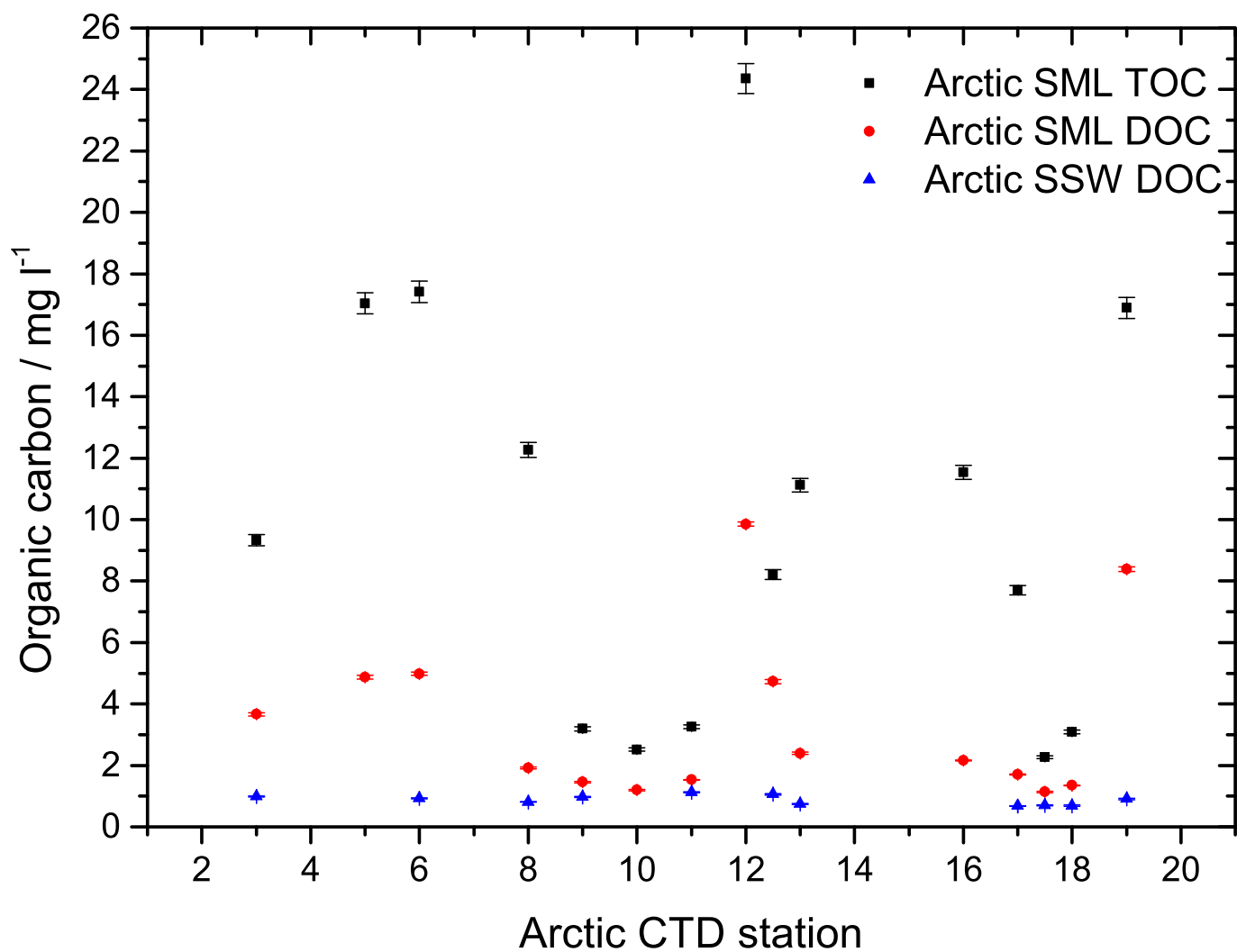


**Extended Data Figure 5 | Correlation of TEP and DOC with the UT-CFDC  $RH_{ice}$  onsets.** **a, b,** Ice nucleation  $RH_{ice}$  onsets for Northeast Pacific (see Extended Data Table 1) samples plotted against measured DOC concentration (**a**) and TEP enrichment factor (**b**). Error bars represent the experimental uncertainty in relative humidity with respect to ice in the UT-CFDC.



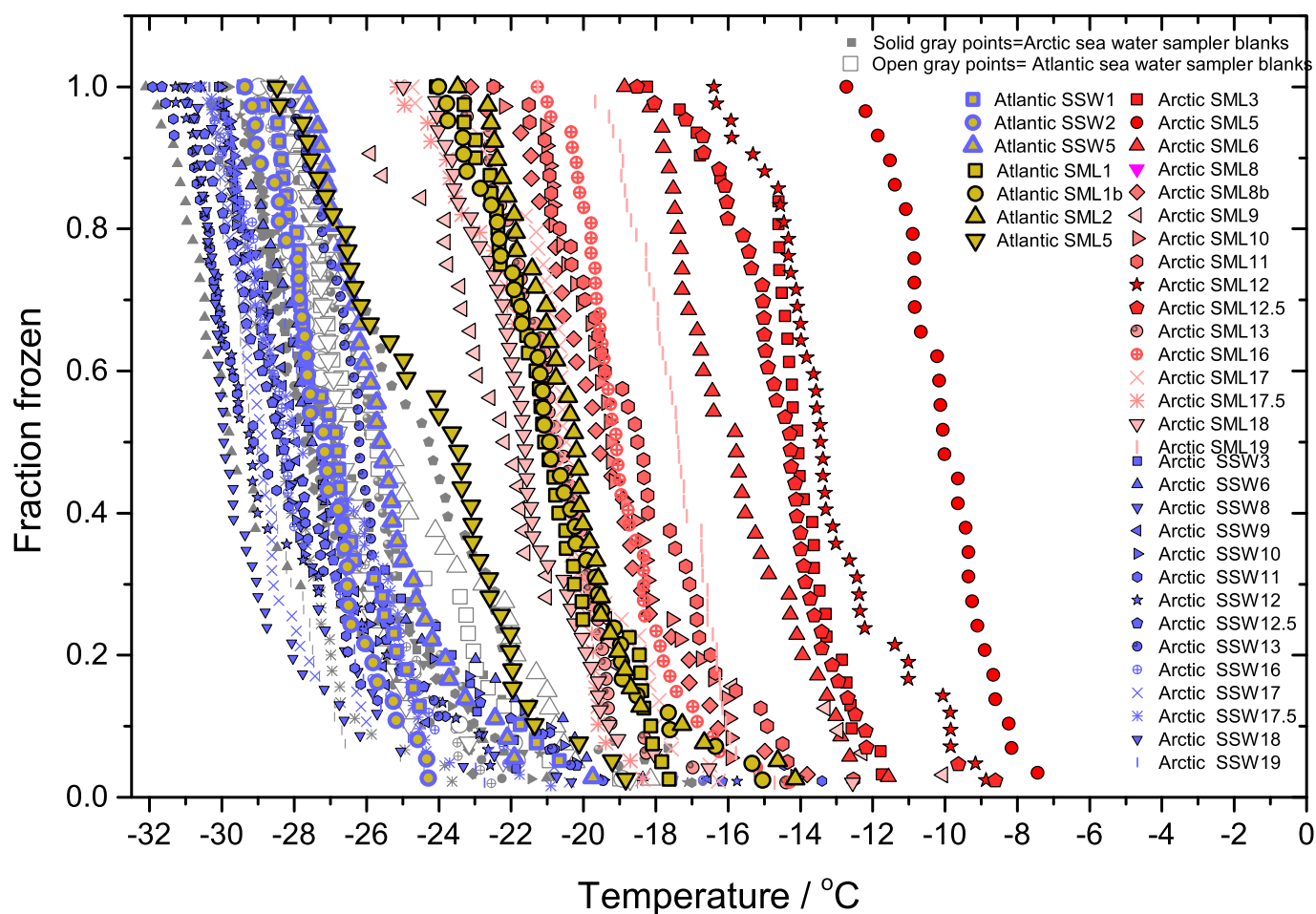
**Extended Data Figure 6 | Ice-nucleating activity of diatom cells and their exudates.** WACIFE frozen fraction curves derived from 60–129- $\mu\text{m}$ -sized droplets ( $\sim 0.4$  nl volume) as a function of temperature. Green symbols indicate diatom exudates in 0.1- $\mu\text{m}$ -filtered sea water. Blue and red symbols represent 0.1- $\mu\text{m}$ -filtered sea water devoid of exudates with and without the addition of growth media, respectively. All temperatures have been corrected for freezing point depression to pure water conditions from their initial aqueous solution water activity,  $a_w = 0.985$  (open circles), 0.97 (open squares),

0.96 (filled diamonds), 0.95 (open diamonds), 0.94 (filled circles), 0.925 (open triangle), 0.90 (asterisks). Shaded areas illustrate ranges of observed heterogeneous ice nucleation of intact and fragmented diatom cells (green) and homogeneous ice nucleation of aqueous NaCl droplets (grey) for similar  $a_w$  values<sup>6,32</sup>. Error bars represent the instrumental uncertainty in temperature measurement. Predicted homogeneous freezing temperatures for similar sized water droplets are indicated by the grey bar<sup>31,40</sup>.



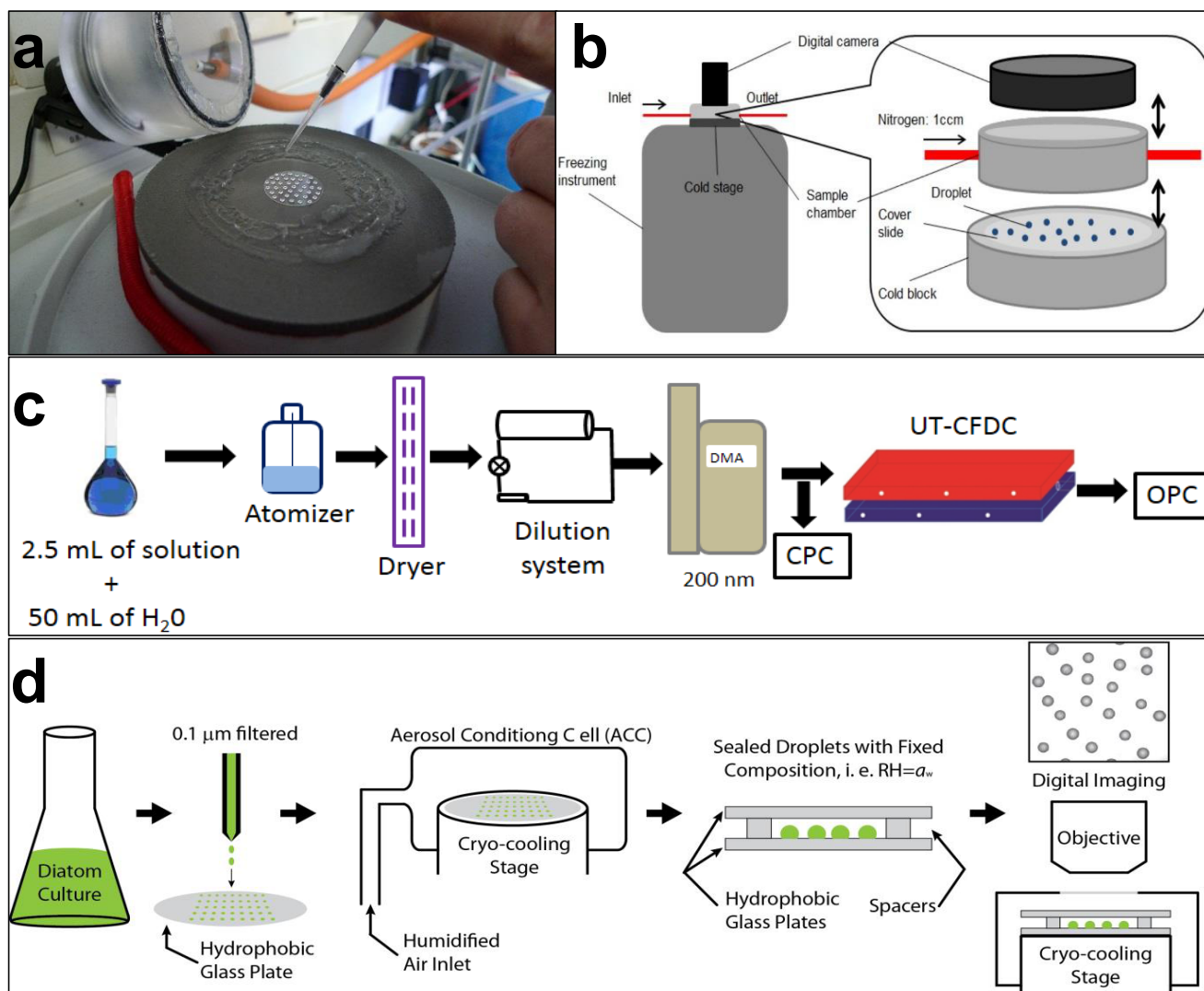
**Extended Data Figure 7 | TOC and DOC measurements for Arctic samples.** Arctic SML TOC and DOC measurements and Arctic SSW DOC measurements. TOC error bars represent the measured 2% coefficient of variance. DOC sample error was calculated as the coefficient of variation from the mean and standard deviation of three sample replicates. For comparison here we provide the Atlantic TOC measurements; Atlantic SML1 =  $5.954 \pm 0.185 \text{ mg l}^{-1}$ , Atlantic SML2 =  $4.643 \pm 0.135 \text{ mg l}^{-1}$ .





**Extended Data Figure 8** |  $\mu\text{l}$ -NIP1 freezing curves for Arctic and Atlantic samples uncorrected for freezing depression caused by salts. Fraction frozen curves for 1  $\mu\text{l}$  droplet freezing experiments using Arctic and Atlantic ocean samples, uncorrected for freezing point depression. SML, SSW and boat

flushing water (grey points; symbols correspond to those for SML sampled at the same locations) to check for the absence of contaminant INPs before sampling.



#### Extended Data Figure 9 | Summary of ice nucleation experimental setups.

**a**, Pipetting 1  $\mu\text{L}$  droplets onto a hydrophobic glass slide. **b**, Schematic of the  $\mu\text{L}$ -NIP cold stage used for immersion mode droplet freezing experiments. **c**, Schematic of the experimental setup for cirrus cloud relevant experiments. CPC, condensation particle counter; DMA, differential mobility analyser; OPC, optical particle counter. **d**, Schematic of the water-activity-controlled

immersion freezing experiment (WACIFE) for freezing of micrometre-sized droplets containing diatom exudates as a function of water activity,  $a_w$ , and relative humidity, RH. Images are not to scale. The procedure for preparing and freezing droplets of filtered and autoclaved natural sea water with and without added f/2 nutrients droplets is similar except 0.1  $\mu\text{m}$  filtration is not required.

Extended Data Table 1 | Details of the sampled SMLs and SSWs

Date	Time / UTC	Location (see Extended data Fig. 1)	Longitude	Latitude	Samples collected	
					SML	SSW/sampling depth
ACCACIA						
19/07/2013	08:22	3	20°41.700W	70°13.494N	Arctic SML3	Arctic SSW3 / 2 m
21/07/2013	07:56	5	20°18.336W	71°53.692N	Arctic SML5	Arctic SSW5 / 5 m
22/07/2013	08:00	6	13°06.120W	73°06.340N	Arctic SML6	Arctic SSW6 / 2 m
24/07/2013	08:02	8	08°43.511W	75°02.956N	Arctic SML8b	Arctic SSW8 / 2 m
25/07/2013	08:07	9	15°31.311W	76°31.902N	Arctic SML9	Arctic SSW9 / 2 m
26/07/2013	14:24	10	5°18.642W	76°16.141N	Arctic SML10	Arctic SSW10 / 2 m
27/07/2013	09:19	11	7°04.645W	77°57.427N	Arctic SML11	Arctic SSW11 / 2 m
28/07/2013	06:15	12	7°01.993W	78°53.663N	Arctic SML12	Arctic SSW12 / 2 m
29/07/2013	08:50	12.5	5°13.61W	77°27.207N	Arctic SML12.5	Arctic SSW12.5 / 2 m
30/07/2013	12:01	13	7°35.043E	74°48.828N	Arctic SML13	Arctic SSW13 / 2 m
03/08/2013	10:20	16	23°56.620E	80°08.900N	Arctic SML16	Arctic SSW16 / 2 m
04/08/2013	07:58	17	33°43.967E	83°18.630N	Arctic SML17	Arctic SSW17 / 0.4 m
04/08/2013	20:48	17.5	33°43.231E	83°18.381N	Arctic SML17.5	Arctic SSW17.5 / 2 m
05/08/2013	07:53	18	26°07.684E	82°41.5N	Arctic SML18	Arctic SSW18 / 2 m
06/08/2013	07:59	19	34°49.928E	81°00.1N	Arctic SML 19	Arctic SSW19 / 0.3 m
WACS II						
23/05/2014	16:13	1	62.3256W	40.41335N	Atlantic SML1	Atlantic SSW1 / 5m
23/05/2014	18:20	1b	62.195878W	40.399563N	Atlantic SML1b	
26/05/2014	18:00	2	61.672607W	42.356586N	Atlantic SML2	Atlantic SSW2 / 5m
04/06/2014	20:30	5	70.5252W	40.68428N	Atlantic SML5	Atlantic SSW5 / 0.5m
NETCARE						
12/08/2013		BC coast	125° 54W	48°93N	BC coast SML1	BC coast SSW1 / 0.5m
12/08/2013		BC coast	125° 54W	48°93N	BC coast SML2	BC coast SSW2 / 0.5m
15/08/2013		BC coast	125°33W	48°54N	BC coast SML3	
15/08/2013		BC coast	125°33W	48°54N	BC coast SML4	
14/06/2013		NE Pacific	138°40.0W	49°34.0N	Pacific SML1	
16/06/2013		NE Pacific	145°00.0W	50°00.0N	Pacific SML2	
16/06/2013		NE Pacific	145°00.0W	50°00.0N	Pacific SML3	Pacific SSW3 / 1 m

Details are given of samples collected during the ACCACIA Arctic cruise, WACS II Atlantic cruise and the NETCARE project. Northeast (NE) Pacific samples were collected as part of the Line P time series, cruise 2013–2017. British Columbia (BC) coast samples were collected in Terrace Bay on the western coast of Vancouver Island (Canada) and at a location approximately 3 km offshore from Ucluelet. Location numbers relate to the maps shown in Extended Data Fig. 1.



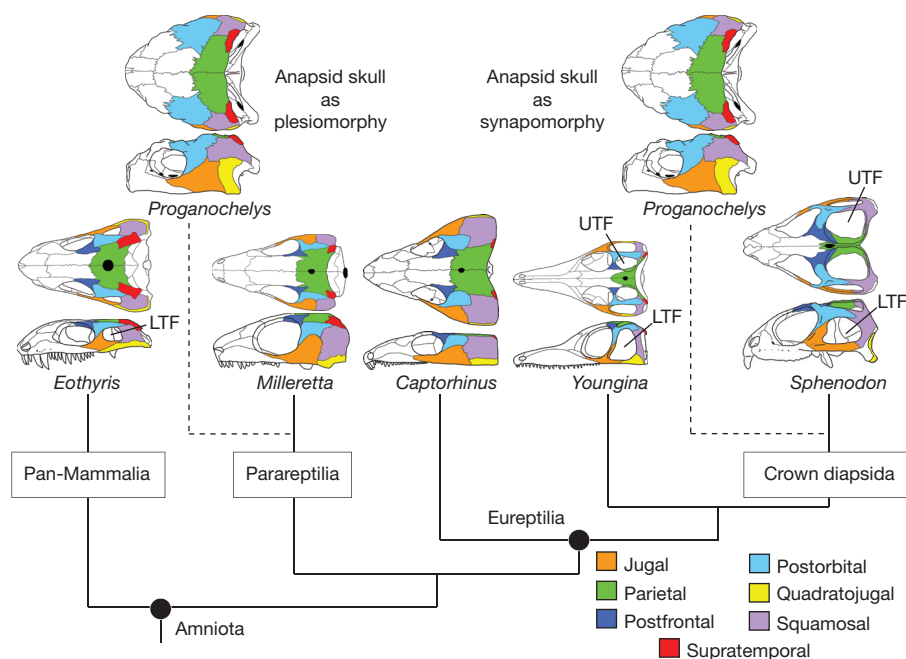
# Evolutionary origin of the turtle skull

G. S. Bever<sup>1,2,3</sup>, Tyler R. Lyson<sup>3,4</sup>, Daniel J. Field<sup>5</sup> & Bhart-Anjan S. Bhullar<sup>5,6</sup>

Transitional fossils informing the origin of turtles are among the most sought-after discoveries in palaeontology<sup>1–5</sup>. Despite strong genomic evidence indicating that turtles evolved from within the diapsid radiation (which includes all other living reptiles<sup>6,7</sup>), evidence of the inferred transformation between an ancestral turtle with an open, diapsid skull to the closed, anapsid condition of modern turtles remains elusive. Here we use high-resolution computed tomography and a novel character/taxon matrix to study the skull of *Eunotosaurus africanus*, a 260-million-year-old fossil reptile from the Karoo Basin of South Africa, whose distinctive postcranial skeleton shares many unique features with the shelled body plan of turtles<sup>2–4</sup>. Scepticism regarding the status of *Eunotosaurus* as the earliest stem turtle arises from the possibility that these shell-related features are the products of evolutionary convergence. Our phylogenetic analyses indicate strong cranial support for *Eunotosaurus* as a critical transitional form in turtle evolution, thus fortifying a 40-million-year extension to the turtle stem and moving the ecological context of its origin back onto land<sup>8,9</sup>. Furthermore, we find unexpected evidence that *Eunotosaurus* is a diapsid reptile in the process of becoming secondarily anapsid. This is important because categorizing the skull based on the number of openings in the complex of dermal bone covering the adductor chamber has long held sway in amniote systematics<sup>10</sup>, and still represents a common organizational scheme for teaching the evolutionary history of the group. These discoveries allow us to articulate a detailed and testable hypothesis of fenestral closure along the turtle stem. Our results suggest that *Eunotosaurus*

represents a crucially important link in a chain that will eventually lead to consilience in reptile systematics, paving the way for synthetic studies of amniote evolution and development.

At least 270 million years<sup>6</sup> of evolution within the reptile crown clade has produced a panoply of cranial forms. From the hyperkinetic anatomy of snakes to the encephalized and highly visual architecture of birds, the reptile skull is an increasingly popular model for understanding the evolution and development of vertebrate adaptation<sup>11,12</sup>. Turtles are an important yet enigmatic piece of this evolutionary puzzle. The earliest uncontroversial stem turtles<sup>1,13</sup> exhibit an anapsid skull with an adductor chamber concealed by bone (Fig. 1). Although emargination has modified this dermal covering in the vast majority of crown-group turtles<sup>14</sup>, the absence of temporal fenestration is a feature shared by currently recognized crown- and stem-group turtles, the immediate fossil outgroups of the amniote crown clade, and many early reptiles (sauropsids)<sup>10</sup>. If this absence reflects conservation of the ancestral amniote condition, then turtles are an extant remnant of an early reptile radiation that excludes the other living forms (tuatara, lizards, snakes, crocodilians, birds). If, however, turtles are nested within the radiation of anatomically diapsid reptiles, which includes both the diapsid crown group and a majority of its stem lineage<sup>15</sup>, then the anapsid skull of turtles is a secondary configuration built on an ancestrally diapsid structural plan. Despite the strong support that a diapsid origin of turtles enjoys from genomic data sets<sup>6,7</sup>, no direct palaeontological evidence yet exists for the loss of a diapsid skull along the turtle stem. This situation epitomizes a general discord between the fossil record and the molecular signature of living taxa that



**Figure 1 | Competing hypotheses for the origin of the anapsid skull of modern turtles.** Historically, this closed condition was accepted as the conservation of the ancestral amniote state with turtles originating among early, long-extinct forms<sup>10,15</sup>. More recent analyses largely reject this hypothesis for a turtle origin within crown-group Diapsida<sup>16,17</sup>—a radiation that includes modern lizards, snakes, tuatara, crocodilians, and birds, and that is generally characterized by a skull with upper and lower temporal fenestrae (UTF and LTF, respectively).

<sup>1</sup>Department of Anatomy, New York Institute of Technology, College of Osteopathic Medicine, Old Westbury, New York 11568, USA. <sup>2</sup>Division of Paleontology, American Museum of Natural History, New York, New York 10024, USA. <sup>3</sup>Evolutionary Studies Institute, University of the Witwatersrand, Private Bag 3, P.O. WITS, Johannesburg 2050, South Africa. <sup>4</sup>Department of Earth Sciences, Denver Museum of Nature and Science, Denver, Colorado 80205, USA. <sup>5</sup>Department of Geology & Geophysics and Peabody Museum of Natural History, Yale University, New Haven, Connecticut 06520, USA. <sup>6</sup>Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA.



Figs 1 and 3) that corresponds to a rare combination of ossified orbital cartilages present at least in *Proganochelys quenstedti* among stem turtles<sup>21</sup>.

These data resolve the cranial–postcranial conflict in that both modules now support *Eunotosaurus* as an early stem turtle, thus bolstering evidence for the turtle stem in the terrestrial ecosystem of Gondwana approximately 260 million years ago (Supplementary Tables 1–3). Considering that *Eunotosaurus* is widely accepted as lying outside the pandiapsid radiation, this shared signal seems only to exacerbate the phylogenetic gap between the phenotypic and genotypic explanations of turtle origins. Although it is the case that neither our parsimony nor Bayesian analyses recover the dominant molecular solution of a unique turtle–archosaur relationship, both approaches do agree that turtles arose somewhere within the greater diapsid radiation (Extended Data Figs 4 and 5). Our morphological results remain ambiguous as to whether that origin is nested within the diapsid crown clade or among those stem forms expressing a morphologically diapsid skull. The lack of clear morphological support for the refined position of turtles within Eureptilia probably reflects some combination of the rate at which these stem lineages diversified and our poor understanding of their respective fossil records<sup>22</sup>. For example, the early lepidosaur stem comprises comparatively few taxa, and the most conservative stem archosaurs bear a striking resemblance to apparent proximal stem diapsids, throwing into question the sequence of acquisition and degree of variability of crown diapsid autapomorphies<sup>23</sup>.

The presence of a lower temporal fenestra in *Eunotosaurus* supports the hypothesis that the characteristically closed cheek of modern turtles is a secondary condition that evolved through expansion of the jugal, quadratojugal, and squamosal<sup>24</sup>. Singular expression of a lower temporal fenestra once unambiguously diagnosed mammals and their stem lineage<sup>10</sup> but is now recognized in a number of phylogenetically disparate ‘anapsid’ parareptiles<sup>25</sup>, and may represent the ancestral condition for the amniote crown<sup>26</sup>. This observed pattern of concentrated homoplasy near the evolutionary origin of a character state is congruent with the concept of a protracted zone of variability that may commonly confound attempts to resolve the early history of clades and character systems<sup>27</sup>.

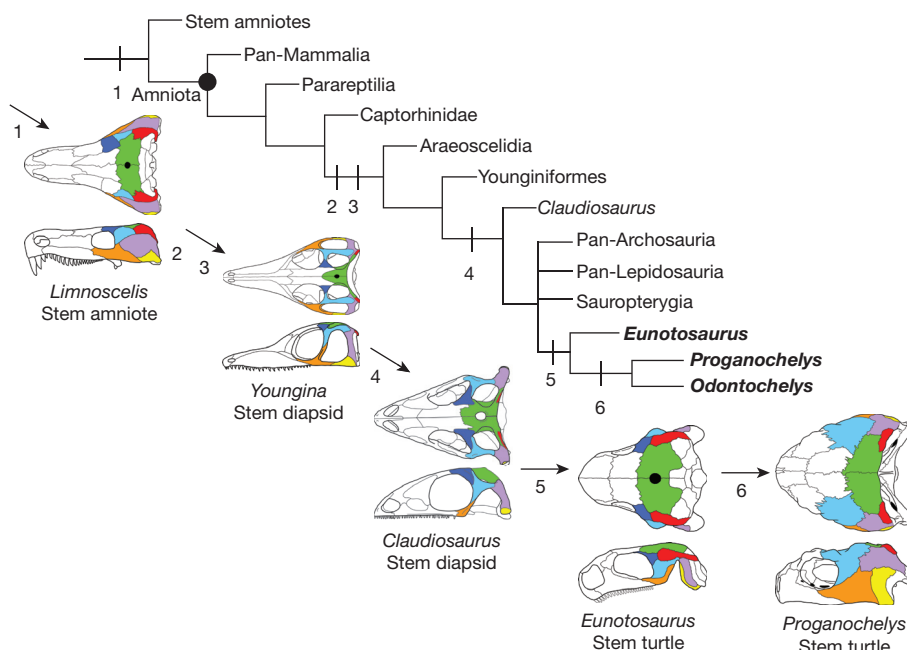
The amniote upper temporal fenestra has enjoyed a famously uncomplicated history, being a nearly consistent fixture of the diapsid body plan since its first appearance in the Carboniferous<sup>28</sup>. The combined morphologies expressed in the juvenile and adult specimens of

*Eunotosaurus* provide not only the earliest direct evidence of an upper temporal fenestra in a putative stem turtle, but the first evidence for how that fenestra may have closed before the origin of the turtle crown clade. This evidence supports a model of temporal closure whose initial steps include a significant expansion of the dermal roof (postorbital, squamosal, and probably parietal) that first constricts and then closes the upper temporal fenestra (Fig. 4). Such steps are expected components of any hypothesis where turtles evolve from a diapsid ancestor. The addition of *Eunotosaurus* to this model is significant because it provides the first empirical evidence of these transitional expansions, which in turn allows the timing of these transformations to be anchored within the geological history of diapsids. *Eunotosaurus* also provides insight into the possible role of developmental timing in producing the modern anapsid condition. Phylogenetic acceleration (peramorphosis) of the inferred postnatal trajectory of dermal expansion around the temporal fenestrae of *Eunotosaurus* may explain the transition from a *Eunotosaurus*-like morphology to that expressed in *Odontochelys*, *Proganochelys*, and more crown-ward turtles.

The role of the supratemporal is an aspect of the *Eunotosaurus* model of temporal closure that would not have been predicted based on earlier studies. An anteriorly expanded supratemporal that develops late in postnatal development to cover the upper temporal fenestra must currently be considered an autapomorphy of *Eunotosaurus*. Future fossil discoveries will determine whether an expanded supratemporal was ancestral to crown-ward turtles, but it is important to stress that an expanded supratemporal is not a necessary component of an evolutionary model of fenestral closure in turtles that has *Eunotosaurus* as a central figure. For example, an analogous secondary expansion of the supratemporal partially or completely covers the upper temporal fenestra of the marine thalattosaurs<sup>29</sup>. Moreover, the construction of the modern turtle skull played out over a time span of tens of millions of years, and it is well attested that dermal bones in major vertebrate lineages can shift back and forth considerably relative to the underlying tissues on these timescales; for instance in the complex history of the synapsid skull and the shoulder girdle of sarcopterygians (and tetrapods, most notably turtles themselves<sup>30</sup>).

It is thus evident that the turtle skull, like the turtle postcranium, underwent profound modifications during its history that similarly obscured anatomical evidence for phylogenetic affinities by the time the crown-group condition was reached. The ecological context in which the earliest stem turtles lost their upper temporal fenestra is

**Figure 4 | Generalized amniote phylogeny showing sequence of major transformations in the origin of the turtle skull.** (1) Ancestral crown amniote retains anapsid condition of amniote stem. (2) and (3) LTF and UTF appear, producing a fully diapsid skull. (4) Loss of lower temporal bar results in ventrally open LTF. (5) Size of UTF and LTF constricted through postnatal expansion of surrounding dermal elements (most notably postorbital and squamosal). Superficial covering of the UTF by the supratemporal, as expressed in *Eunotosaurus*, may or may not be ancestral to modern turtles. (6) Closure of LTF and UTF, perhaps involving a peramorphic shift in dermal bone growth. A secondary reduction of the supratemporal also may have occurred at this point.





unclear, although the upper fenestra in extant diapsids relates to the bulging of the pseudotemporalis muscle mass. It is thus likely that fenestral closure along the turtle stem had implications for the masticatory apparatus, and these implications may be reflected in the various modifications of the rostral portion of the skull.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 28 April; accepted 3 July 2015.**

**Published online 2 September 2015.**

- Li, C., Wu, X.-C., Rieppel, O., Wang, L.-T. & Zhao, L.-J. An ancestral turtle from the Late Triassic of southwestern China. *Nature* **456**, 497–501 (2008).
- Lyson, T. R., Bever, G. S., Bhullar, B.-A. S., Joyce, W. G. & Gauthier, J. A. Transitional fossils and the origin of turtles. *Biol. Lett.* **6**, 830–833 (2010).
- Lyson, T. R., Bever, G. S., Scheyer, T. M., Hsiang, A. Y. & Gauthier, J. A. Evolutionary origin of the turtle shell. *Curr. Biol.* **23**, 1113–1119 (2013).
- Lyson, T. R. *et al.* Origin of the unique ventilatory apparatus of turtles. *Nat. Commun.* **5**, 5211 (2014).
- Schoch, R. R. & Sues, H.-D. A Middle Triassic stem-turtle and the evolution of the turtle body plan. *Nature* **523**, 584–587 (2015).
- Wang, Z. *et al.* The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nature Genet.* **45**, 701–706 (2013).
- Field, D. J. *et al.* Toward concision in reptile phylogeny: miRNAs support an archosaur, not lepidosaur, affinity for turtles. *Evol. Dev.* **16**, 189–196 (2014).
- Joyce, W. G. & Gauthier, J. A. Palaeoecology of Triassic stem turtles sheds new light on turtle origins. *Proc. R. Soc. Lond. B* **271**, 1–5 (2004).
- Scheyer, T. M. & Sander, P. M. Shell bone histology indicates terrestrial palaeoecology of basal turtles. *Proc. R. Soc. Lond. B* **274**, 1885–1893 (2007).
- Gauthier, J., Kluge, A. G. & Rowe, T. Amniote phylogeny and the importance of fossils. *Cladistics* **4**, 105–209 (1988).
- Cundall, D. & Irish, F. in *Biology of the Reptilia* Vol. 20 (eds Gans, C., Gaunt, A. S., Adler, K.) 349–692 (Society for the Study of Amphibians and Reptiles, 2008).
- Bhullar, B.-A. S. *et al.* Birds have paedomorphic dinosaur skulls. *Nature* **487**, 223–226 (2013).
- Gaffney, E. S. The comparative osteology of the Triassic turtle *Proganochelys*. *Bull. Am. Mus. Nat. Hist.* **194**, 1–263 (1990).
- Werneberg, I. Temporal bone arrangements in turtles: an overview. *J. Exp. Zool. B* **318**, 235–249 (2012).
- Müller, J. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G., Wilson, M. V. H., Wilson, R., Cloutier, R.) 379–408 (Verlag Dr. Friedrich Pfeil, 2004).
- Lee, M. S. Y. Turtle origins: insights from phylogenetic retrofitting and molecular scaffolds. *J. Evol. Biol.* **26**, 2729–2738 (2013).
- Day, M., Rubidge, B., Almond, J. & Sifelani, J. Biostratigraphic correlation in the Karoo: the case of the Middle Permian parareptile *Eunotosaurus*: research letter. *S. Afr. J. Sci.* **109**, 1–4 (2013).
- Cox, C. B. The problematic Permian reptile *Eunotosaurus*. *Bull. Br. Mus. Nat. Hist.* **18**, 165–196 (1969).
- Keyser, A. W. & Gow, C. E. First complete skull of the Permian reptile *Eunotosaurus africanus* Seeley. *S. Afr. J. Sci.* **77**, 417–420 (1981).
- Gow, C. E. A reassessment of *Eunotosaurus africanus* Seeley (Amniota: Parareptilia). *Palaeont. Afr.* **34**, 33–42 (1997).
- Bhullar, B.-A. S. & Bever, G. S. An archosaur-like laterosphenoid in early turtles (Reptilia: Pantestudines). *Breviora* **518**, 1–11 (2009).
- Reisz, R. R., Modesto, S. P. & Scott, D. M. A new Early Permian reptile and its significance in early diapsid evolution. *Proc. R. Soc. Lond. B* **278**, 3731–3737 (2011).
- Bickelmann, C., Müller, J. & Reisz, R. R. The enigmatic diapsid *Acerosodontosaurus piveteaui* (Reptilia: Neodiapsida) from the Upper Permian of Madagascar and the paraphyly of “younginiform” reptiles. *Can. J. Earth Sci.* **46**, 651–661 (2009).
- Müller, J. Early loss and multiple return of the lower temporal arcade in diapsid reptiles. *Naturwissenschaften* **90**, 473–476 (2003).
- Tsuji, L. A. & Müller, J. Assembling the history of the Parareptilia: phylogeny, diversification, and a new definition of the clade. *Fossil Rec.* **12**, 71–81 (2009).
- Piñero, G., Ferigolo, J., Ramos, A. & Laurin, M. Cranial morphology of the Early Permian mesosaurid *Mesosaurus tenuidens* and the evolution of the lower temporal fenestration reassessed. *C. R. Palevol* **11**, 379–391 (2012).
- Bever, G. S., Gauthier, J. A. & Wagner, G. P. Finding the frame shift: digit loss, developmental variability, and the origin of the avian hand. *Evol. Dev.* **13**, 269–279 (2011).
- Reisz, R. R. *A Diapsid Reptile from the Pennsylvanian of Kansas* (Univ. of Kansas, 1981).
- Rieppel, O. *Clarazia* and *Hescheleria*: a re-investigation of two problematical reptiles from the Middle Triassic of Monte San Giorgio (Switzerland). *Palaeontogr. Abt. A* **195**, 101–129 (1987).
- Lyson, T. R. *et al.* Homology of the enigmatic nuchal bone reveals novel reorganization of the shoulder girdle in the evolution of the turtle shell. *Evol. Dev.* **15**, 1–9 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Botha-Brink, E. Butler, S. Kaal, E. De Kock, J. Neveling and R. Smith for access to *Eunotosaurus* specimens. M. Fox and Z. Erasmus prepared fossil material. M. Colbert, J. Maisano, M. Hill and J. Thostenson are acknowledged for their help with the digital data. We thank A. Balanoff, D. Dykes, J. Gauthier, R. Hill, B. Rubidge, R. Smith and K. de Queiroz for helpful discussions. G.S.B. extends special thanks to the Academic Technologies Group at NYIT for their support in the digital visualization of anatomical data.

**Author Contributions** G.S.B. designed the study, processed the CT data, performed the analytical work, constructed the figures, and wrote the paper. T.R.L. performed analytical work, assisted writing the paper, and assisted with figures. D.J.F. and B.-A.S.B. performed analytical work and assisted writing the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.S.B. (gbever@nyit.edu).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Specimens and CT scanning.** Although *Eunotosaurus africanus* is known from a surprisingly large number of specimens ( $n > 44$ ), relatively few of these include cranial material (see below). Our study was built largely around the adult morphology of CM777 and CM86-341, and the juvenile features of SAM-PK-K7909. NMQR3299 was also examined, but relatively poor preservation restricted its contribution to our cranial assessments. The skull of CM 777 was scanned at the University of Texas High-Resolution X-ray CT Facility (UTCT). Scanning was performed using no filter, an air wedge, a voltage of 200 kV and a current of 0.17 mA. The resulting images were then processed for the removal of ring artefacts. The specimen was scanned along the coronal axis for a total of 1,003 slices with an image resolution of  $1024 \times 1024$  pixels and a reconstructed field of view of 34 mm. Voxel size (mm) is  $0.03657 \times 0.03320 \times 0.03320$ . Each image has a reconstructed field of view of 34 mm. Additional images are available at DigiMorph ([http://www.digimorph.org/specimens/Eunotosaurus\\_africanus](http://www.digimorph.org/specimens/Eunotosaurus_africanus)). Original slice data are available on request.

NMQR3299 (skull and postcranial skeleton) was scanned at UTCT with no filter, an air wedge, a voltage of 200 kV, and a current of 0.19 mA. Ring artefacts were removed. Scanning was performed along the coronal axis for a total of 1,764 slices with a resolution of  $1024 \times 1024$  pixels and a reconstructed field of view of 62 mm. Voxel size (mm) is  $0.06065 \times 0.06065 \times 0.06618$ . No digital segmentation of this data set was performed.

CM86-341 was scanned at the American Museum of Natural History Microscopy and Imaging Facility (AMNH MIF) using a copper filter, an air wedge, a voltage of 150 kV, and a current of 124 mA. The specimen was scanned along the coronal axis for a total of 950 slices with a resolution of  $1024 \times 1024$  pixels and a reconstructed field of view of 35 mm. Voxel size (mm) is  $0.03403 \times 0.03403 \times 0.03403$ . Digital segmentation of the recognizable cranial elements was performed using VGStudioMax2.1.

This study also includes novel morphological data derived from the recent physical preparation of two specimens of *Eunotosaurus* (using small PaleoTools microjack and microscope). CM86-341 was prepared in 2010–2012 by M. Fox (Peabody Museum, Yale University) under the direction of J. Gauthier. SAM-PK-K7670 was prepared in 2014–2015 by Z. Erasmus (Iziko Museums of South Africa) under the direction of R. Smith.

**List of examined cranial specimens of *Eunotosaurus africanus*.** CM86-341: beautifully preserved partial skull, completely articulated neck with a few cervical ribs, and complete carapace (nine dorsal vertebrae and nine pairs of dorsal ribs) (Fig. 2 and Extended Data Fig. 1). CM777: articulated skull, neck, elongate cervical ribs, shoulder girdle, limb elements, and cranial half of carapace including dorsal vertebrae and ribs (Fig. 2 and Extended Data Figs 2 and 3). Unnumbered CM specimen figured in ref. 19. NMQR3299: mostly complete skeleton including articulated skull (Fig. 2). NMQR3474: impression of a mostly articulated skeleton, including the skull. SAM-PK-K7670: highly weathered nodule with mostly complete skeleton including cranial two-thirds dorsal ribs and vertebrae, impressions of the cervical vertebrae, and an impression of the skull. SAM-PK-K7909: weathered nodule with complete shell and shoulder girdle, articulated neck, and essentially complete and articulated skull (Fig. 3). Juvenile based on its small size and expression of numerous features indicative of skeletal immaturity in reptiles (that is, unfused scapula and coracoid)<sup>31,32</sup>.

Institutional abbreviations used are as follows: CM, Council for Geosciences, Pretoria; NMQR, National Museum, Bloemfontein; RC, Rubidge Collection, Graaff-Reinet; SAM-PK, Iziko Museums of South Africa, Cape Town; YPM, Peabody Museum, Yale University, New Haven.

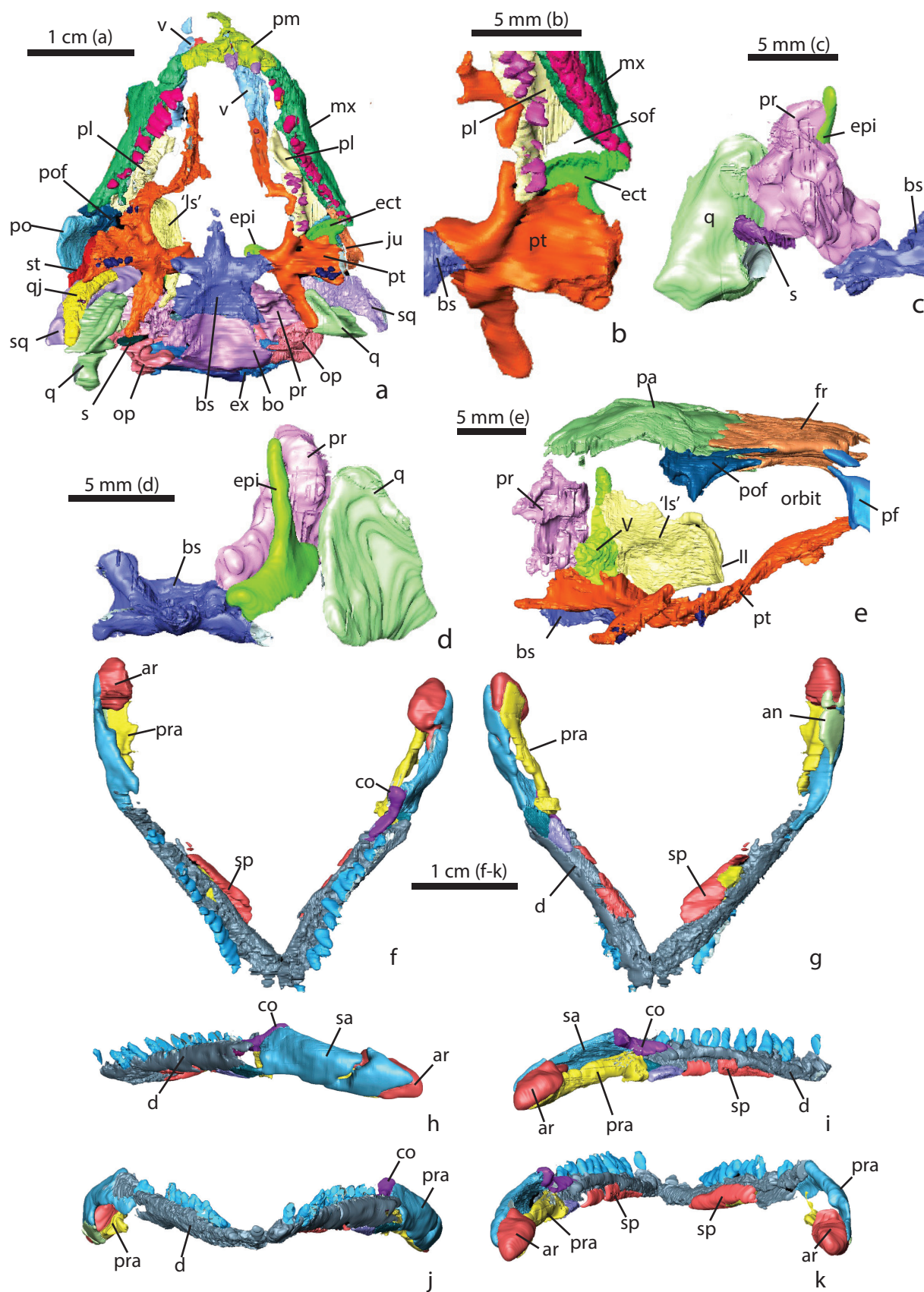
**Phylogenetic analysis.** The phylogenetic relationships of *Eunotosaurus africanus* were assessed using a novel character/taxon matrix consisting of 268 characters (174 cranial, 94 postcranial) scored for 47 taxa distributed broadly across amniotes with emphasis on Pan-Reptilia. Our primary purpose was to determine how an enhanced understanding of the cranial anatomy of *Eunotosaurus* affects its phylogenetic status as an early stem turtle and the topological position of turtles in general. This is especially important considering that many of the more compelling *Eunotosaurus*–turtle synapomorphies are features associated with the iconic, shelled body plan<sup>2–4</sup>. While of considerable interest, the shared expression of these features also raises the question of convergence—perhaps there are only so many ways to build a shell. The skull serves as a relatively independent module on which to test the *Eunotosaurus*–turtle hypothesis.

Tree topologies were inferred using both maximum parsimony and Bayesian optimality criteria. To specifically assess the dominant phylogenetic signal within the cranial data, separate analyses were performed on the complete matrix and a matrix restricted to cranial characters. The maximum parsimony tree was generated using TNT1.1<sup>33</sup>. *Seymouria* spp. was specified as the outgroup, and heuristic searches were conducted using tree-bisection-reconnection (TBR) branch swapping with 1,000 replicates of random stepwise sequence addition. Minimum branch lengths were set to collapse. Support for each node was measured by calculating Bremer support and bootstrap frequencies, with 1,000 bootstrap replicates and 1,000 random sequence addition replicates. Characters 34, 62, 67, 85, 107, 118, 148, 155, 193 and 220 (Supplementary Table 1) were treated as ordered as their derived states were interpreted as non-mutually exclusive (that is, the possession of either derived state reflects shared information that should be considered in the analysis). The implications of this approach were tested by comparing the results with iterations where all characters were analysed as unordered.

Bayesian phylogenetic analyses were run using MrBayes (v3.2.2)<sup>34</sup> on the CIPRES Science Gateway<sup>35</sup>. The Mk model<sup>36</sup> was used to analyse the full and cranial-only character matrices with gamma-distributed rate variation and variable coding. All analyses were performed with a sampling frequency of 1,000, two concurrent runs, and four Metropolis-coupled chains ( $T = 0.1$ ) for 30,000,000 generations. Characters 34, 62, 67, 85, 107, 118, 148, 155, 193 and 220 were again treated as ordered. All analyses were checked for convergence using standard MrBayes diagnostics (for example, PRSF < 0.01, mixing between chains > 20%) and Tracer (v1.5)<sup>37</sup> (for example, ESS > 200). A 25% relative burn-in was implemented for all summary statistics.

A list of the cranial characters and their definitions is provided as Supplementary Table 1. Postcranial characters are taken directly from ref. 3. Supplementary Table 2 provides character scores and Supplementary Table 3 provides a list of synapomorphies for each of the reptile clades within which *Eunotosaurus* is nested. Supplementary Table 4 lists the observed specimens and primary references from which we compiled our character matrix.

31. Romer, A. S. *Osteology of the Reptiles* (Univ. Chicago Press, 1956).
32. Maisano, J. A. Terminal fusions of skeletal elements as indicators of maturity in squamate reptiles. *J. Vertebr. Paleontol.* **22**, 268–275 (2002).
33. Goloboff, P. A., Farris, J. & Nixon, K. TNT: a free program for phylogenetic analysis. *Cladistics* **24**, 774–786 (2008).
34. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
35. Miller, M. A., Pfeiffer, W. & Schwartz, T. In *Proceedings of the Gateway Computing Environments Workshop* 1–8 (IEEE, 2010).
36. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).
37. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).



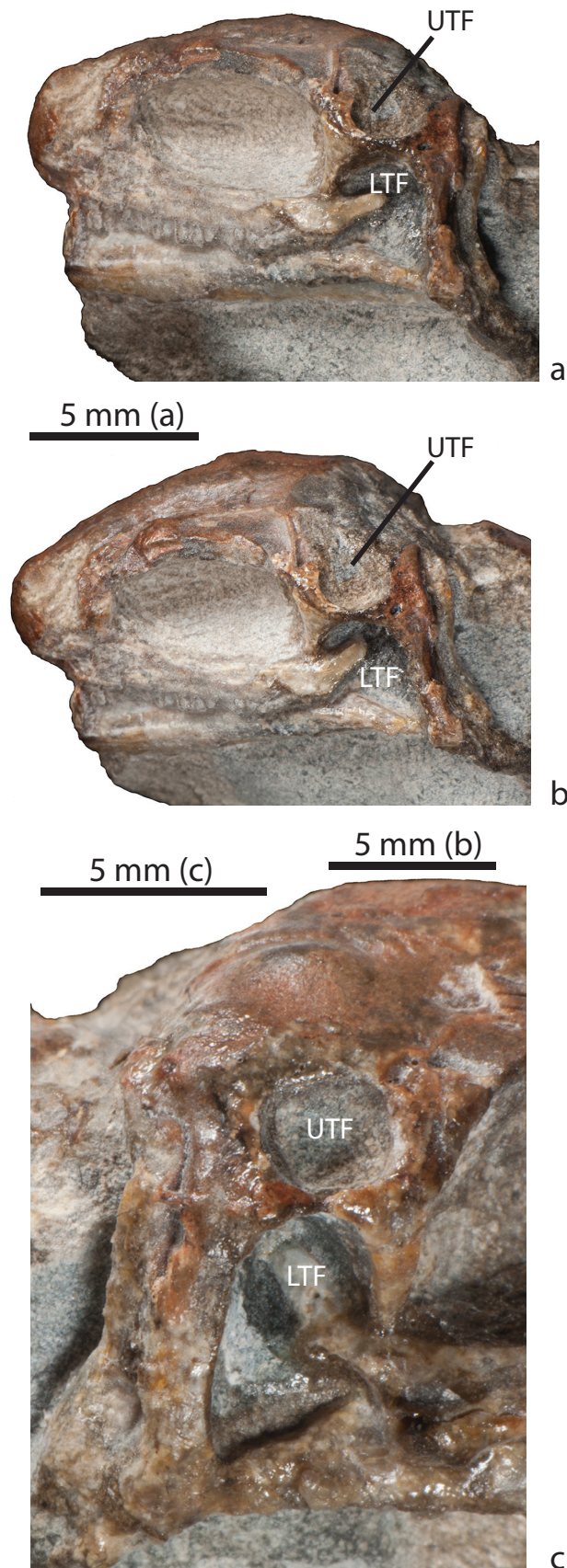


**Extended Data Figure 1 | Digital reconstruction of segmented cranial**

**elements of *Eunotosaurus africanus* CM 777.** **a**, Palatal view with the lower jaws digitally removed and major roofing elements not rendered.

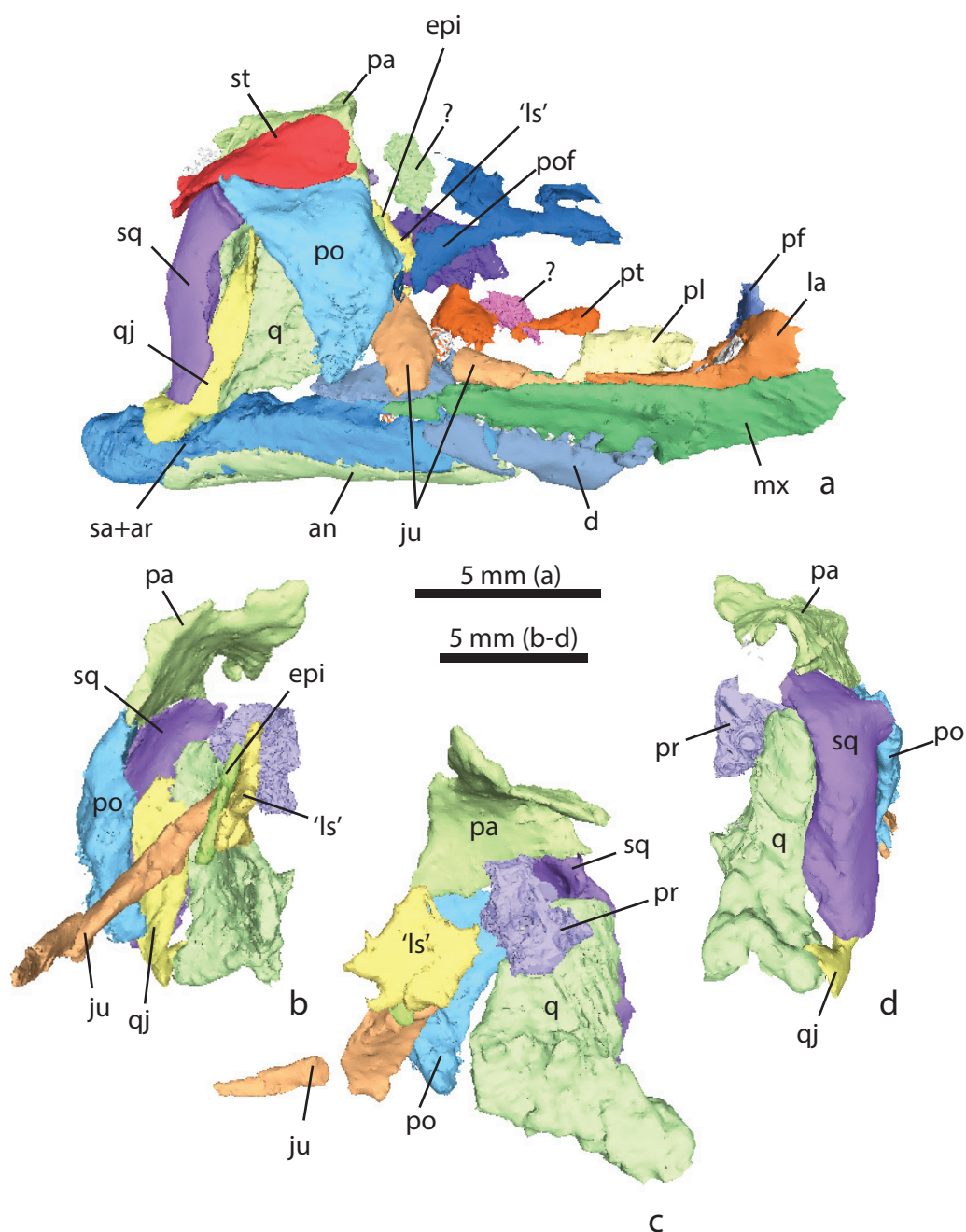
**b**, Anteromedial view of left palate showing moderately sized suborbital fenestra. **c**, **d**, Posteromedial (**c**) and anterolateral (**d**) views of left quadrate, prootic, stapes, epipterygoid and midline parabasisphenoid. **e**, Right lateral view of anterior braincase wall and surrounding elements. Note sutural contact of prootic and quadrate. **f–k**, Lower jaws in dorsal (**f**), ventral (**g**), left lateral (**h**),

medial (left jaw) (**i**), anterior (**j**), and posterior (**k**) views. an, angular; ar, articular; bs, parabasisphenoid; co, coronoid; d, dentary; ect, ectopterygoid; epi, epipterygoid; fr, frontal; ju, jugal; la, lacrimal; ls, laterosphenoid; mx, maxilla; op, opisthotic; pa, parietal; pf, prefrontal; pl, palatine; pm, premaxilla; po, postorbital; pof, postfrontal; pr, prootic; pra, prearticular; pt, pterygoid; q, quadrate; qj, quadratojugal; s, stapes; sq, squamosal; sof, suborbital fenestra; sp., splenial; st, supratemporal; sa, surangular; v, vomer; IL, inferred exit point for orbital nerve; V, prootic incisure, exit point for trigeminal nerve.



**Extended Data Figure 2 | The juvenile skull of *Eunotosaurus africanus* (SAM-PK-K7909) showing clear expression of both LTF and UTF. a, b, Left lateral view with the rostrum held horizontally (a) and slightly downturned (b). c, Close-up view of fenestrated cheek in right lateral view. The size of the**

fenestrae decreases in the late stages of postnatal ontogeny through expansion of the surrounding dermal bones. The upper temporal fenestra is eventually obscured by the late-stage ontogenetic development of an elongate supratemporal.

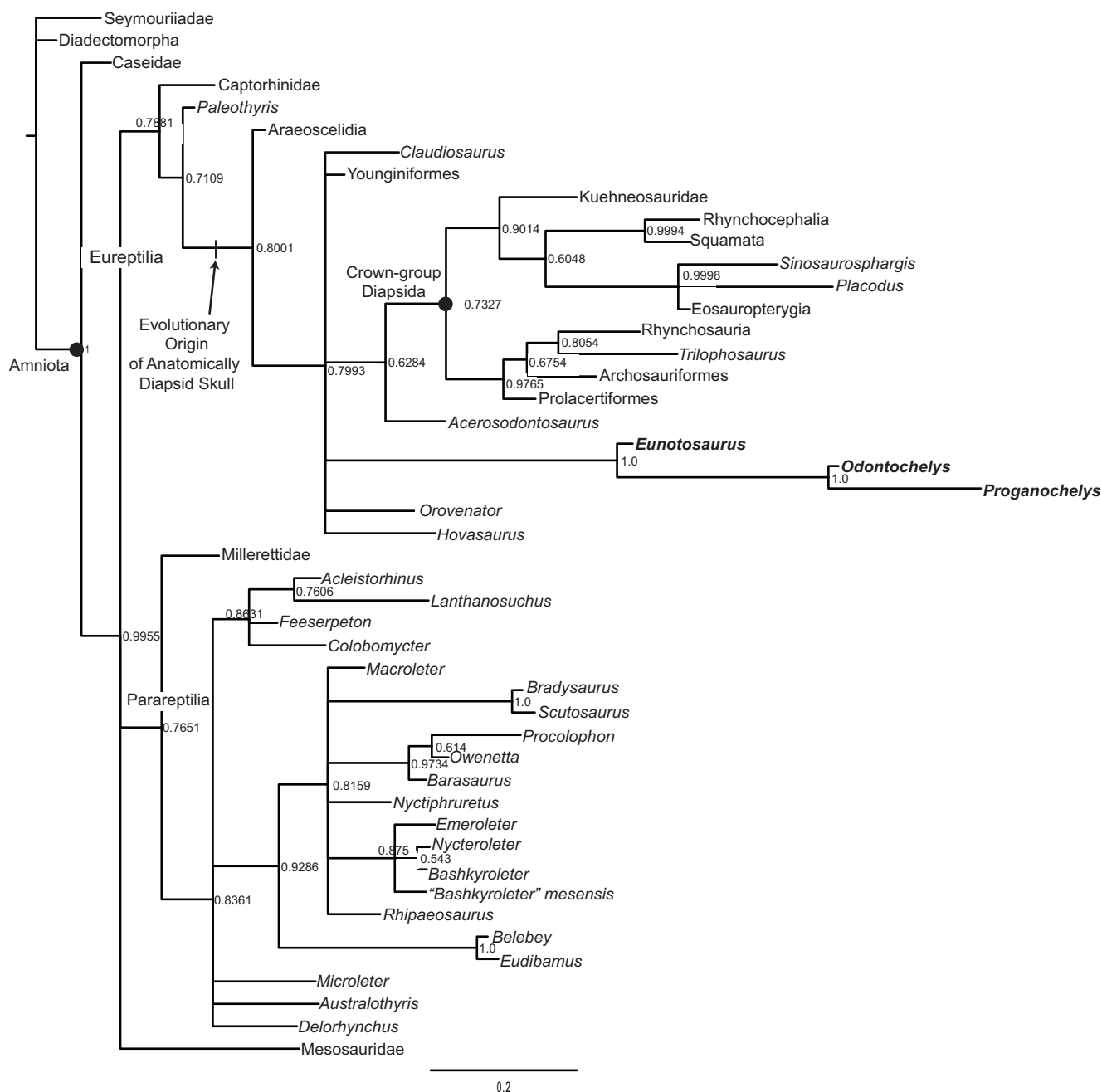


**Extended Data Figure 3 | Digitally segmented and reconstructed cranial elements of *Eunotosaurus africanus* (CM86-341).** a–d, Right lateral view (a), anterior (b), posterior (c), and right medial (d) views. an, angular; d, dentary; epi, epipterygoid; ju, jugal; la, lacrimal; 'ls', 'laterosphenoid'; mx,

maxilla; pa, parietal; pf, prefrontal; pl, palatine; po, postorbital; pof, postfrontal; pt, pterygoid; q, quadrate; qj, quadratojugal; sa + ar, surangular and articular; sq, squamosal; st, supratemporal; UTF, upper temporal fenestra; ?, unclear identity.

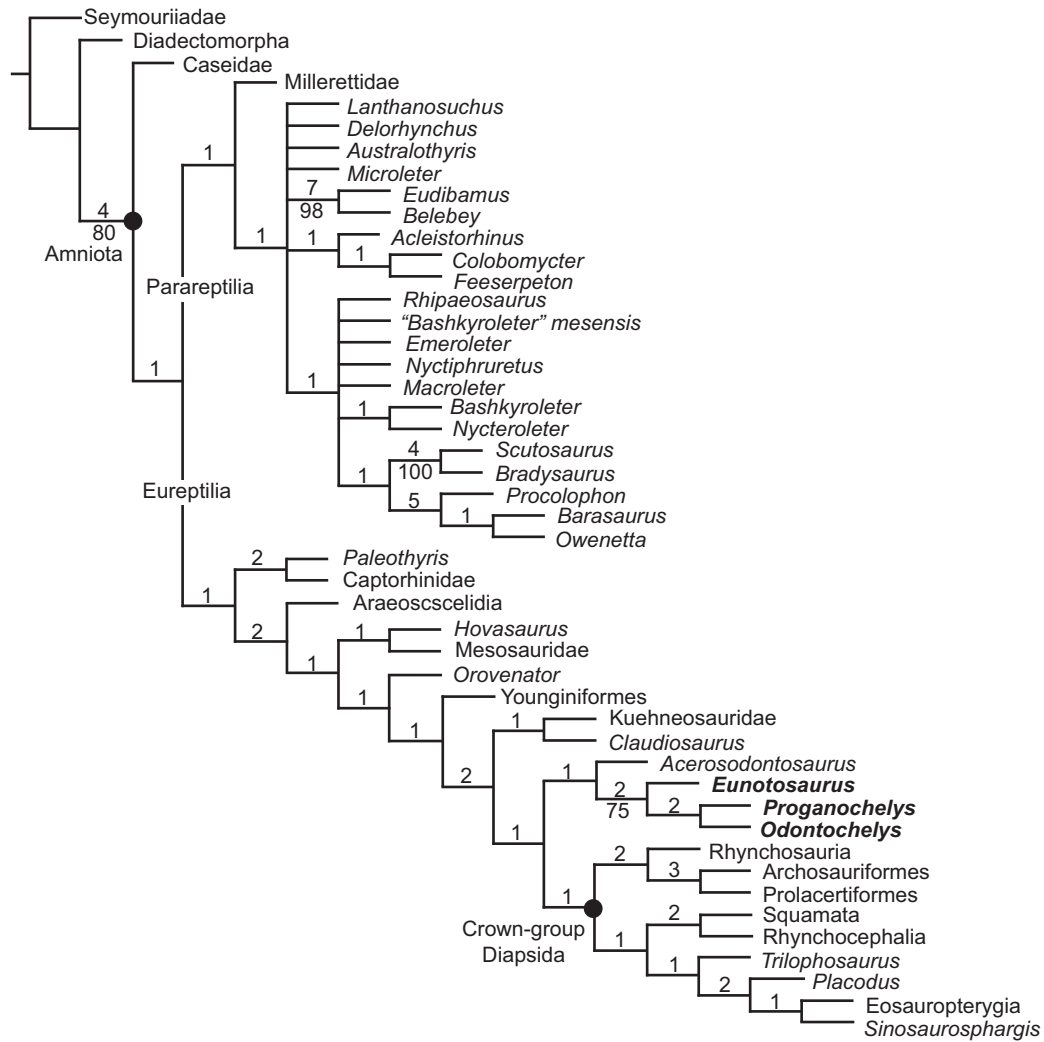






**Extended Data Figure 5 | Bayesian tree topology derived from total matrix (50% majority rule consensus).** An exclusive *Eunotosaurus*–turtle clade is recovered with 100% posterior probability. This pan-turtle lineage is nested within the radiation of anatomically diapsid reptiles; however, in contrast to the parsimony solution, turtles are excluded from crown-group Diapsida. The

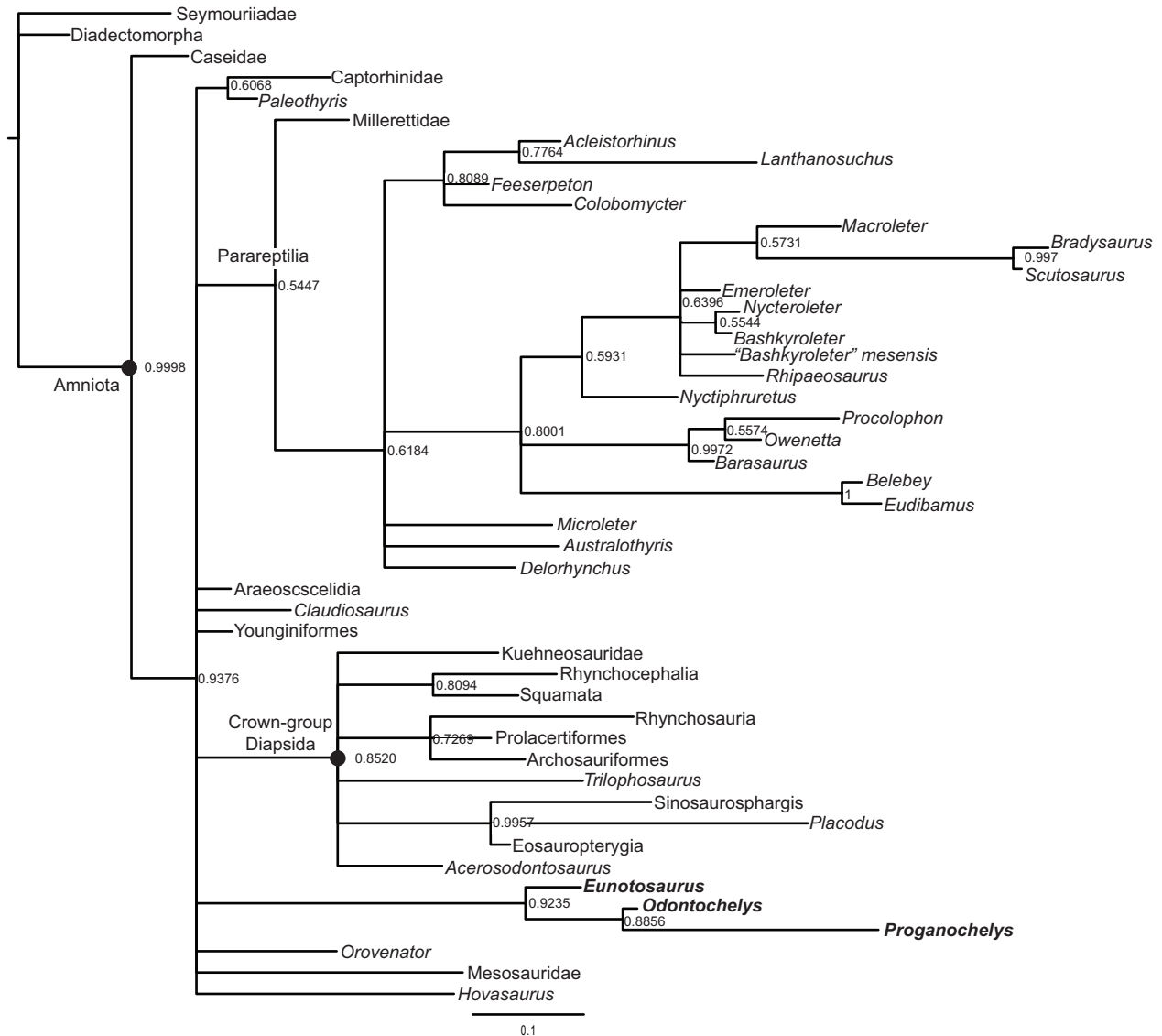
Bayesian results agree with the parsimony in revealing strong support that: (1) *Eunotosaurus* is an early stem-group turtle; and (2) the ancestral stem turtle expressed a fully diapsid skull. The two analyses also agree that there is currently poor morphological support for a refined position of turtles within the greater diapsid radiation.



**Extended Data Figure 6 | Strict consensus of 13 most parsimonious trees recovered from cranial-only matrix.** The *Eunosaurus*–turtle clade is recovered, which supports the hypothesis that the postcranial synapomorphies

of *Eunosaurus* and turtles are homologous and not the products of convergence. Tree length = 777; consistency index = 0.3956; retention index = 0.4743.





**Extended Data Figure 7 | Bayesian tree topology derived from cranial-only matrix (50% majority rule consensus).** When studied in isolation, cranial anatomy provides poor resolution of the deep divergences within Pan-Reptilia, but a *Eunotosaurus*–turtle signal is clearly present.

# Arithmetic and local circuitry underlying dopamine prediction errors

Neir Eshel<sup>1</sup>, Michael Bukwich<sup>1</sup>, Vinod Rao<sup>1</sup>, Vivian Hemmelder<sup>1</sup>, Ju Tian<sup>1</sup> & Naoshige Uchida<sup>1</sup>

**Dopamine neurons are thought to facilitate learning by comparing actual and expected reward<sup>1,2</sup>. Despite two decades of investigation, little is known about how this comparison is made. To determine how dopamine neurons calculate prediction error, we combined optogenetic manipulations with extracellular recordings in the ventral tegmental area while mice engaged in classical conditioning. Here we demonstrate, by manipulating the temporal expectation of reward, that dopamine neurons perform subtraction, a computation that is ideal for reinforcement learning but rarely observed in the brain. Furthermore, selectively exciting and inhibiting neighbouring GABA ( $\gamma$ -aminobutyric acid) neurons in the ventral tegmental area reveals that these neurons are a source of subtraction: they inhibit dopamine neurons when reward is expected, causally contributing to prediction-error calculations. Finally, bilaterally stimulating ventral tegmental area GABA neurons dramatically reduces anticipatory licking to conditioned odours, consistent with an important role for these neurons in reinforcement learning. Together, our results uncover the arithmetic and local circuitry underlying dopamine prediction errors.**

Associative learning depends on comparing predictions with outcomes<sup>3,4</sup>. When outcomes match predictions, learning is not required. When outcomes violate predictions, animals must update their predictions to reflect experience. Dopamine neurons are thought to promote this process by encoding reward prediction error, or the difference between the reward an animal receives and the reward it expected to receive<sup>1,2</sup> (see Supplementary Information).

Despite extensive study, how dopamine neurons calculate prediction error remains largely unknown. Theories of reinforcement learning predict that dopamine neurons perform subtraction, simply calculating actual reward minus predicted reward (or, in temporal difference models, the value of the current state minus the value of the previous state)<sup>1</sup>. However, dopamine neurons could also perform division, an equally fundamental and arguably more common neural computation<sup>5</sup>. The arithmetic underlying prediction errors has never been investigated.

To probe how dopamine neurons calculate prediction error, we recorded from the ventral tegmental area (VTA) (Extended Data Figs 1a and 2a–c) while mice ( $n = 5$ ) performed a classical conditioning task with two interleaved trial types (Fig. 1a). In roughly half of the trials, we delivered reward unexpectedly, in the absence of any cue. In these trials, both the timing and size of reward were unexpected. In the other half of the trials, an odour cue predicted the timing of reward, but the size was still unexpected. By comparing responses to these two trial types, we could determine how temporal expectation modulates individual dopamine neurons across a range of firing rates. The light-gated ion channel, channelrhodopsin (ChR2), was expressed selectively in dopamine neurons, enabling us to identify neurons as dopaminergic on the basis of their responses to light<sup>6</sup> (Extended Data Fig. 3a–g).

Consistent with previous results<sup>6,7</sup>, dopamine neurons increased their responses with increasing reward size (example neuron in Extended Data Fig. 4a). Much like sensory neurons in response to stimuli of increasing intensity, dopamine neurons showed a gradual,

monotonic response, well-fit by a saturating Hill function (orange trace in Fig. 1c; note that VTA GABA neurons do not show the same monotonic response, Extended Data Fig. 5).

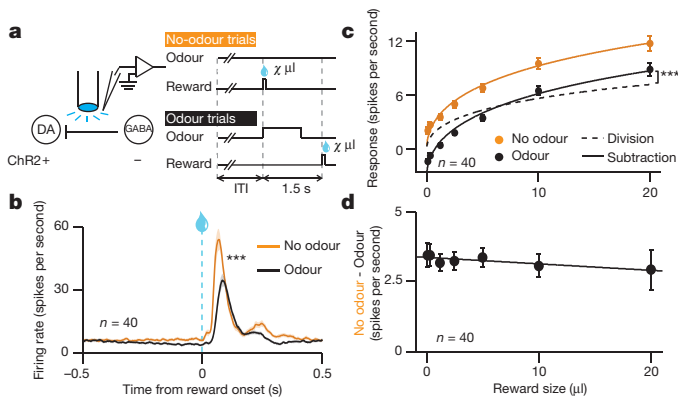
When reward was temporally expected, the responses of dopamine neurons were suppressed ( $P < 0.001$ ,  $t$ -test; example neuron in Extended Data Fig. 4a; population in Fig. 1b). To determine the nature of the suppression, we performed two complementary analyses. First, we fitted dopamine responses both with subtractive and with divisive models (Extended Data Fig. 6a). We found that subtraction was a significantly better fit ( $P < 0.001$ , bootstrap; Fig. 1c and Extended Data Fig. 4b). Second, we plotted the effect of temporal expectation across reward sizes and measured the slope. A divisive process would produce a positive slope, as division should have a larger effect on larger dopamine responses. In contrast, subtraction would produce a slope near zero. We found the latter; regardless of reward size, the odour cue simply shifted the dose–response curve by a constant amount ( $P > 0.05$ , linear regression, Fig. 1d). This subtractive pattern held not just for the population, but also for 35 out of 40 individual neurons (Extended Data Fig. 4c). Thus, consistent with classic theories of reinforcement learning, dopamine neurons appear to be performing subtraction (specifically, output subtraction<sup>8</sup>; see Extended Data Fig. 6a).

Having established the computation, we next wished to determine the input that dopamine neurons subtract. A variety of biological models have been proposed to explain the neural circuit required to calculate prediction errors. Some of these models have situated the calculation at the level of the dopamine neurons<sup>9,10</sup>, while others have suggested that the calculation happens upstream, for instance in the lateral habenula<sup>11,12</sup>, and is then relayed to dopamine neurons. Recently, we demonstrated that GABAergic neurons in the VTA encode reward expectation, showing sustained responses that vary with the timing and size of expected reward<sup>6</sup>. Although these neurons are known to synapse onto nearby dopamine neurons<sup>13</sup> and appear to play a role in conditioned behaviour<sup>14,15</sup>, there has been no direct evidence that dopamine neurons use the VTA GABA signal for prediction-error calculations. Furthermore, although some models of prediction-error calculations call for a ramping expectation function<sup>9,16,17</sup>, which resembles VTA GABA activity, others call for phasic, precisely timed expectation signals<sup>18–20</sup>. Our study allows us to distinguish between these possibilities.

Since we know the normal firing patterns of VTA GABA neurons during classical conditioning<sup>6</sup>, our strategy was to mimic this firing and determine whether it induces subtraction of dopamine neuron responses. In a separate set of mice ( $n = 5$ ), ChR2 was expressed selectively in VTA GABA neurons, enabling us to stimulate these neurons while recording from putative dopamine neurons (Extended Data Figs 1b and 2d–f). Much like the previous task, we unexpectedly delivered rewards of various size (Fig. 2a). In half of the trials, reward was delivered alone; in the other half, reward was delivered during 40 Hz VTA GABA stimulation.

First we confirmed that ChR2 stimulation efficiently excited VTA GABA neurons ( $P < 0.001$ , paired  $t$ -test), adding about ten spikes per second to the baseline firing rate of the neurons (example neuron in Extended Data Fig. 4d; population in Fig. 2b). This laser-evoked

<sup>1</sup>Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.



**Figure 1 | Expectation triggers subtraction of dopamine neuron responses.** **a**, Dopamine (DA) identification recording setup (left) and task (right). ITI, inter-trial interval. **b**, Dopamine neuron firing rates (mean  $\pm$  s.e.m. across neurons) for unexpected (orange) or temporally expected (black) reward. \*\*\* $P$  < 0.001,  $t$ -test. **c**, Dopamine neuron responses (mean  $\pm$  s.e.m.) to different reward sizes. Orange line, fit for unexpected reward. Dotted black line, divisive transformation. Solid black line, subtractive transformation. Subtraction was a better fit (\*\*\* $P$  < 0.001, bootstrap; see Methods and Extended Data Fig. 6e). **d**, Difference between unexpected and expected reward responses (mean  $\pm$  s.e.m.) as a function of reward size.

activity roughly resembled the normal activity of VTA GABA neurons during classical conditioning (Extended Data Fig. 2f).

Next we assessed how VTA GABA stimulation affected putative dopamine neuron responses to reward. As expected, GABA stimulation significantly suppressed dopamine reward responses ( $P$  < 0.001,  $t$ -test; example neuron in Extended Data Fig. 4e; population in Fig. 2c). This suppression could not be fully explained by a shift in baseline activity (Extended Data Fig. 7a–d). Moreover, the dopamine suppression was not due to an association between blue light and reward, as laser delivery failed to elicit expectation-related licking behaviour (Extended Data Fig. 8b). Indeed, a separate group of control mice ( $n$  = 2) expressing green fluorescent protein (GFP) rather than Chr2 in GABA neurons (Extended Data Fig. 2g–i) showed no effect of laser stimulation ( $P$  = 0.78, Fig. 2e, f).

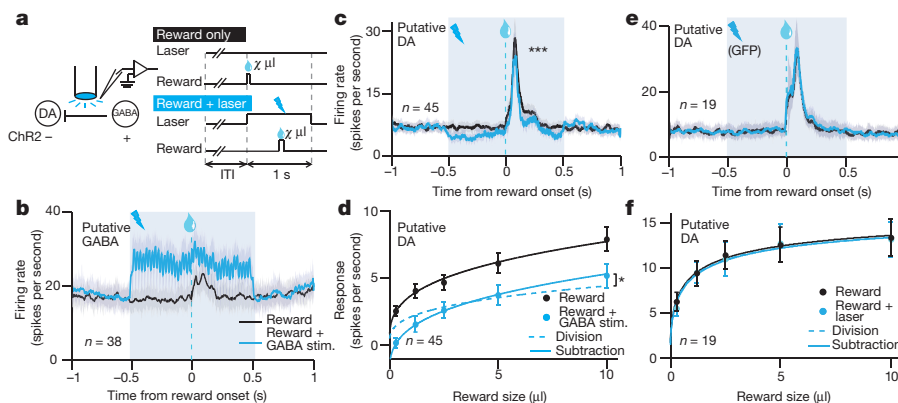
We confirmed that stimulating VTA GABA neurons suppresses phasic dopamine activity, but what is the shape of this suppression? As in our previous experiment, we determined a dopamine dose-response curve and fitted both subtractive and divisive models. We found that the effect of VTA GABA stimulation was subtractive ( $P$  < 0.05, bootstrap; Fig. 2d and Extended Data Fig. 6f, g). This subtractive effect held even when correcting for the baseline-lowering effect of GABA

stimulation ( $P$  < 0.05; see Methods and Extended Data Fig. 6h, i). We conclude that VTA GABA activation mimics the effect of temporal expectation on putative dopamine neurons.

Although we show that VTA GABA activity can account for expectation-like changes in dopamine responses, this does not demonstrate that VTA GABA neurons normally play such a role. To strengthen the causal link between VTA GABA activity and dopamine prediction-error coding, we inhibited VTA GABA neurons during their normal period of activity and asked whether this disrupts dopamine prediction errors. In a separate group of mice ( $n$  = 7), the light-sensitive inhibitory proton pump archaerhodopsin (ArchT)<sup>21</sup> was expressed selectively in VTA GABA neurons (Extended Data Figs 1c, e–g and 9a–c). Mice were trained in a two-odour classical conditioning task, in which odour A predicted reward with 10% probability and odour B predicted reward with 90% probability (Fig. 3a). In 25% of the trials, we delivered green laser to activate ArchT and inhibit VTA GABA neurons for 1 s around reward outcome.

We first confirmed that laser stimulation significantly suppressed expectation-related activity in putative VTA GABA neurons ( $P$  = 0.001,  $t$ -test; individual neurons in Extended Data Fig. 4f, h; population in Fig. 3b). Next, we assessed how inhibiting VTA GABA neurons modified dopamine activity. Normally, putative dopamine neurons had reduced reward responses when a cue predicted reward delivery ( $P$  < 0.001, paired  $t$ -test; Fig. 3c). Inhibiting VTA GABA neurons partially reversed this expectation-dependent reduction (individual dopamine neurons in Extended Data Fig. 4g, i; population in Fig. 3d). Thus, when VTA GABA neurons are inhibited, dopamine neurons respond as if reward is less expected. This change was specific to phasic reward responses, and not due solely to a shift in baseline activity (Extended Data Figs 7e–h and 10). Combined with our Chr2 experiment, these results suggest that VTA GABA neurons play a causal role in dopamine prediction-error coding. In particular, they help provide the burst-cancelling expectation signal long anticipated by models of reinforcement learning<sup>16,18,22</sup>.

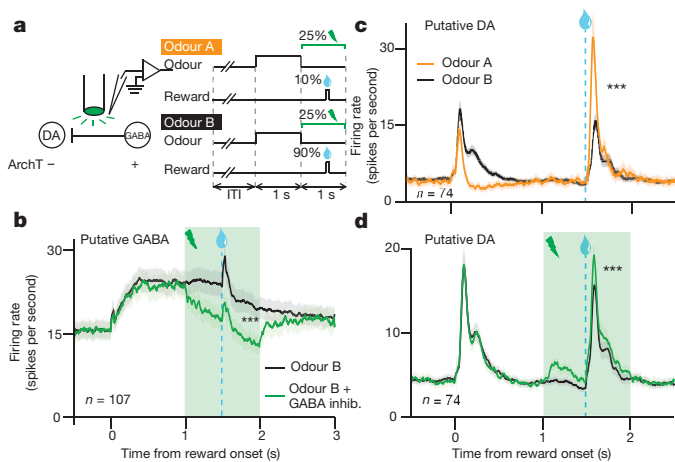
In Figs 2 and 3 we report that VTA GABA excitation and inhibition modulates dopamine prediction-error responses. However, our unilateral optogenetic manipulation did not modify mouse behaviour (Extended Data Fig. 8b, c). To determine whether the VTA GABA expectation signal is important for learning, we designed an additional experiment with bilateral manipulation. In a separate group of mice ( $n$  = 6), Chr2 was expressed selectively in VTA GABA neurons bilaterally. The mice performed a four-odour classical conditioning task, in which odour A was associated with large reward, odours B and D were associated with small reward, and odour C was associated with no reward (Fig. 4a). After training, odour D trials were paired with VTA GABA stimulation. Importantly, the odour–reward associations



**Figure 2 | Selective excitation of VTA GABA neurons mimics the effect of expectation.** **a**, GABA stimulation (stim.) recording setup (left) and task (right). **b**, Firing rate (mean  $\pm$  s.e.m.) of putative VTA GABA neurons with (blue) and without (black) Chr2 stimulation. Light blue box, laser delivery. **c**, Firing rate (mean  $\pm$  s.e.m.) of putative dopamine neurons. \*\*\* $P$  < 0.001,

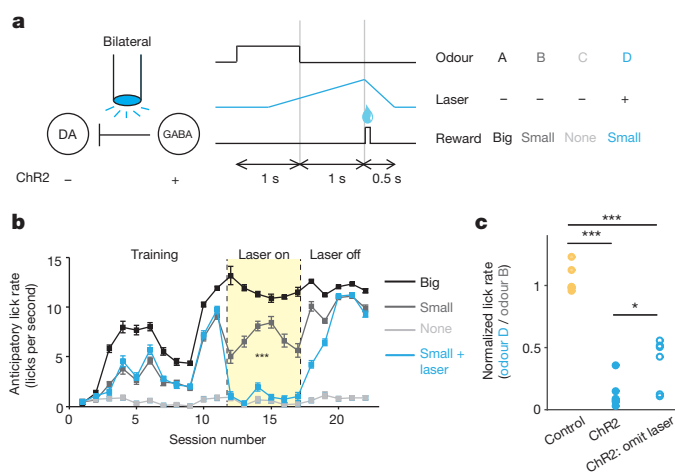
$t$ -test. **d**, Dopamine neuron responses (mean  $\pm$  s.e.m.) to different reward sizes. Black line, fit for unexpected reward. Dotted blue line, divisive transformation. Solid blue line, subtractive transformation. Subtraction was a better fit (\* $P$  < 0.05, bootstrap; see Extended Data Fig. 6g). **e**, **f**, Same as **c** and **d**, respectively, but in GFP-expressing control animals.





**Figure 3 | Selective inhibition of VTA GABA neurons modulates prediction errors.** **a**, GABA inhibition (inhib.) recording setup (left) and task (right). **b**, Firing rate (mean  $\pm$  s.e.m.) of putative VTA GABA neurons during odour B trials with (green) or without (black) laser delivery. \*\*\* $P < 0.001$ , paired  $t$ -test. **c**, Firing rate (mean  $\pm$  s.e.m.) of putative dopamine neurons when reward was delivered after odour A (orange) or odour B (black). \*\*\* $P < 0.001$ , paired  $t$ -test. **d**, Same as **b**, but for putative dopamine neurons. \*\*\* $P < 0.001$ , paired  $t$ -test.

always remained the same. Our hypothesis was that over time, laser stimulation would reduce dopamine prediction-error responses for odour D. As a result, the expected value of odour D should decrease, and mice should lick less for odour D compared with odour B, even though the reward was the same. Indeed, this is what we found: after the laser was introduced, mice licked significantly less for odour D than for odour B ( $P < 0.001$ , laser  $\times$  odour interaction, mixed effects linear model, Fig. 4b and Extended Data Fig. 8d). This reduction did not occur in a separate group of control mice ( $n = 6$ ) that did not express ChR2 (Extended Data Fig. 8e). Although there was probably a direct effect of GABA stimulation on licking behaviour, as previously discovered<sup>14</sup>, this cannot account for the entire difference, because the reduction remained significant on probe trials, where odour D was not



**Figure 4 | Bilateral excitation of VTA GABA neurons disrupts learned association.** **a**, Schematic of optogenetic setup (left) and behavioural task (right). **b**, For a representative mouse (one of six mice injected with ChR2), anticipatory licks during each session (mean  $\pm$  s.e.m. across trials) for odours A (black), B (dark grey), C (light grey), and D (blue). For sessions 12–17 (pale yellow), odour D was paired with laser. \*\*\* $P < 0.001$ , laser  $\times$  odour interaction, mixed effects model. **c**, Ratio of anticipatory licks for odour D versus odour B during laser sessions. Circles, mice injected with ChR2 (blue) or GFP (yellow). Open circles, probe trials, where laser was omitted after odour D. \* $P < 0.05$ ; \*\*\* $P < 0.001$ ; Wilcoxon rank-sum test.

paired with laser stimulation (Fig. 4c). In other words, previous laser trials caused the mice to learn a new, reduced value for odour D, which persisted even in the absence of the laser. In the prediction-error framework, this new value may have been learned through GABA-induced dips in dopamine firing (see Supplementary Information). Consistent with our physiology results, our behavioural findings imply an important role for VTA GABA neurons in prediction-error learning.

Our study provides the first direct evidence for the arithmetic of dopamine prediction errors. Subtraction is an ideal process for prediction-error coding because it maintains a faithful separation between expected and unexpected rewards, even at the extremes of reward size (Extended Data Fig. 6a). Indeed, most, if not all, models of reinforcement learning have used subtraction to compute prediction error. However, although cortical pyramidal neurons appear capable of subtracting GABA input<sup>23–25</sup>, and modelling studies have explored the biophysics of this process<sup>26–28</sup>, surprisingly few examples of subtraction have been observed in natural settings *in vivo*<sup>29,30</sup>. Our finding that reward expectation reduces dopamine reward responses in a purely subtractive manner sheds light on how such a computation can emerge from a network of neurons, and may provide a framework for other prediction-related processes in the brain.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 June 2014; accepted 23 June 2015.

Published online 31 August 2015.

- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
- Bush, R. R. & Mosteller, F. A mathematical model for simple learning. *Psychol. Rev.* **58**, 313–323 (1951).
- Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. & Prokasy, W.) 64–99 (Appleton-Century-Crofts, 1972).
- Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nature Rev. Neurosci.* **13**, 51–62 (2012).
- Cohen, J. Y., Haesler, S., Vogt, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
- Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
- Silver, R. A. Neuronal arithmetic. *Nature Rev. Neurosci.* **11**, 474–489 (2010).
- Houk, J. C., Adams, J. L. & Barto, A. G. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J. C., Davis, J. L. & Beiser, D. G.) 249–270 (MIT Press, 1995).
- Kawato, M. & Samejima, K. Efficient reinforcement learning: computational theories, neuroscience and robotics. *Curr. Opin. Neurobiol.* **17**, 205–212 (2007).
- Matsumoto, M. & Hikosaka, O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447**, 1111–1115 (2007).
- Hong, S., Zhou, T. C., Smith, M., Saleem, K. S. & Hikosaka, O. Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *J. Neurosci.* **31**, 11457–11471 (2011).
- Omelchenko, N. & Sesack, S. R. Ultrastructural analysis of local collaterals of rat ventral tegmental area neurons: GABA phenotype and synapses onto dopamine and GABA cells. *Synapse* **63**, 895–906 (2009).
- Van Zessen, R., Phillips, J. L., Budygin, E. A. & Stuber, G. D. Activation of VTA GABA neurons disrupts reward consumption. *Neuron* **73**, 1184–1194 (2012).
- Tan, K. R. et al. GABA neurons of the VTA drive conditioned place aversion. *Neuron* **73**, 1173–1183 (2012).
- Hazy, T. E., Frank, M. J. & O'Reilly, R. C. Neural mechanisms of acquired phasic dopamine responses in learning. *Neurosci. Biobehav. Rev.* **34**, 701–720 (2010).
- Rivest, F., Kalaska, J. F. & Bengio, Y. Conditioning and time representation in long short-term memory networks. *Biol. Cybern.* **108**, 23–48 (2014).
- Vitay, J. & Hamker, F. H. Timing and expectation of reward: a neuro-computational model of the afferents to the ventral tegmental area. *Front. Neurobot.* **8**, 4 (2014).
- Ludvig, E. A., Sutton, R. S. & Kehoe, E. J. Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Comput.* **20**, 3034–3054 (2008).
- Tan, C. O. & Bullock, D. A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *J. Neurosci.* **28**, 10062–10074 (2008).
- Han, X. et al. A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Front. Syst. Neurosci.* **5**, 18 (2011).
- Fiorillo, C. D., Song, M. R. & Yun, S. R. Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. *J. Neurosci.* **33**, 4710–4725 (2013).

23. Pi, H.-J. *et al.* Cortical interneurons that specialize in disinhibitory control. *Nature* **503**, 521–524 (2013).
24. Wilson, N. R., Runyan, C. A., Wang, F. L. & Sur, M. Division and subtraction by distinct cortical inhibitory networks *in vivo*. *Nature* **488**, 343–348 (2012).
25. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).
26. Murphy, B. K. & Miller, K. D. Multiplicative gain changes are induced by excitation or inhibition alone. *J. Neurosci.* **23**, 10040–10051 (2003).
27. Ayaz, A. & Chance, F. S. Gain Modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *J. Neurophysiol.* **101**, 958–968 (2009).
28. Holt, G. R. & Koch, C. Shunting inhibition does not have a divisive effect on firing rates. *Neural Comput.* **9**, 1001–1013 (1997).
29. Roy, J. E. & Cullen, K. E. Dissociating self-generated from passively applied head motion: neural mechanisms in the vestibular nuclei. *J. Neurosci.* **24**, 2102–2111 (2004).
30. Rust, N. C., Schwartz, O., Movshon, J. A. & Simoncelli, E. P. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* **46**, 945–956 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Andermann, J. Assad, R. Born, J. Buckholtz, P. Glimcher, J. Maunsell, B. Sabatini, W. Schultz, R. Wilson, and members of the Uchida laboratory for comments on the manuscript; S. Haesler for technical expertise and discussions on experimental design; E. Molnar for histology assistance; C. Dulac for sharing resources; K. Deisseroth for the AAV-FLEX-ChR2 construct; and E. Boyden for the AAV-FLEX-ArchT construct. This work was supported by a Sackler Fellowship in Psychobiology (N.E.) and National Institutes of Health grants T32GM007753 (to N.E.), F30MH100729 (to N.E.), R01MH095953 (to N.U.), and R01MH101207 (to N.U.).

**Author Contributions** N.E. and N.U. designed the recording experiments. N.E., V.R., and N.U. designed the behaviour experiment. N.E., M.B., V.R., V.H., and J.T. collected data. N.E., M.B., and V.R. analysed data. N.E. wrote the manuscript with comments from N.U.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.U. ([uchida@mcb.harvard.edu](mailto:uchida@mcb.harvard.edu)).

## METHODS

**Animals.** We used 33 adult male mice, backcrossed for more than 5 generations with C57/BL6J mice, that were heterozygous for Cre recombinase under the control of either the DAT gene (B6.SJL-*Slc6a3<sup>tm1.1(cre)Bkmm</sup>*/J, The Jackson Laboratory)<sup>31</sup> or the *Vgat* gene (*Vgat-ires-Cre*)<sup>32</sup>. We used 5 animals in the dopamine-identification task (Fig. 1), 7 in the GABA stimulation task (Fig. 2), 9 in the GABA inhibition task (Fig. 3), and 12 in the behavioural experiment (Fig. 4). Animals were housed on a 12 h dark/12 h light cycle (dark from 7:00 to 19:00) and performed the task at the same time each day. In the behavioural experiment, animals were randomly assigned to either the experimental or control groups, and the experimenters were blinded to the assignment during all surgeries, behavioural sessions, and individual mouse analyses. All procedures were approved by the Harvard University Institutional Animal Care and Use Committee.

**Surgery and viral injections.** All surgeries were performed under aseptic conditions with animals under either ketamine/medetomidine (60 and 0.5 mg kg<sup>-1</sup>, intraperitoneal, respectively) or isoflurane (1–2% at 0.5–1.0 l min<sup>-1</sup>) anaesthesia. Analgesia (ketoprofen, 5 mg kg<sup>-1</sup> intraperitoneally; buprenorphine, 0.1 mg kg<sup>-1</sup>, intraperitoneally) was administered postoperatively. For the recording experiments, mice underwent two surgeries, both stereotactically targeting left VTA (from bregma: 3.0 mm posterior, 0.8 mm lateral, 4–5 mm ventral). In the first surgery, we injected 200–500 nl adeno-associated virus (AAV) to enable cell-type identification or manipulation (see below). After 2–4 weeks, we performed a second surgery to implant a head plate and microdrive containing six to eight tetrodes and an optical fibre, as described<sup>6</sup>. Recording sites are displayed in Extended Data Fig. 1. For the behavioural experiment, mice underwent a single surgery in which we injected 500 nl AAV into VTA bilaterally, and then implanted a headplate and a dual-optic fibre cannula (300 µm diameter, Doric Lenses) custom-designed to target bilateral VTA.

The viral injections differed among the four experiments. In the dopamine identification experiment (Fig. 1), we injected AAV (serotype 5) carrying an inverted ChR2 (H134R) fused to the fluorescent reporter enhanced yellow fluorescent protein (eYFP) and flanked by double *loxP* sites<sup>33,34</sup>. We previously showed that expression of this virus in dopamine neurons is highly selective and efficient<sup>6</sup>. In both the GABA stimulation experiment (Fig. 2) and the behavioural experiment (Fig. 4), we injected the same AAV-FLEX-ChR2-eYFP construct or, for control mice, we injected AAV5-GFP (University of North Carolina Vector Core). Finally, in the GABA inhibition experiment (Fig. 3), we injected AAV (serotype 1 or 8) carrying an inverted ArchT<sup>21</sup> fused to the fluorescent reporter GFP and flanked by double *loxP* sites (University of North Carolina Vector Core). Expression of ArchT was almost 100% selective to GABA neurons and about 50% efficient, for both AAV1 and AAV8 (Extended Data Fig. 1e–g). In both the ChR2 and ArchT experiments, no virus-expressing cell bodies were observed distant from the injection site (for example, in the striatum or the cortex), implying that the virus was not taken up by axons in the VTA and transported retrogradely to input areas.

**Behavioural tasks.** After more than 1 week of recovery, mice were water-restricted in their cages. Weight was maintained above 90% of baseline body weight. Animals were head-restrained and habituated for 1–2 days before training. Odours were delivered with a custom-made olfactometer<sup>35</sup>. Each odour was dissolved in mineral oil at 1/10 or 1/100 dilution. Thirty microlitres of diluted odour was placed inside a filter-paper housing, and then further diluted with filtered air by 1:20 to produce a 1,000 ml min<sup>-1</sup> total flow rate. Odours included isoamyl acetate, (+)-carvone, 1-hexanol, *p*-cymene, ethyl butyrate, and 1-butanol, and differed for different animals. In the recording experiments, licks were detected by breaks of an infrared beam placed in front of the water tube. In the behavioural experiments, licks were detected by contact with a water tube connected to a capacitive sensing circuit (Teensy, PJRC).

Each trial began with 1 s odour delivery, followed by a delay (either 0.5 s or 1 s), and a reward outcome. In the dopamine identification experiment (Fig. 1), the outcome ranged from 0.1 µl to 20 µl water; in the GABA stimulation experiment (Fig. 2), the outcome ranged from 0.3 µl to 10 µl water; in the GABA inhibition experiment (Fig. 3), the outcome was either 0 µl or 3.75 µl water; and in the behavioural experiment (Fig. 4), the outcome was 0, 2, or 5 µl water. Inter-trial intervals were drawn from an exponential distribution (mean 7.6 s), resulting in a flat hazard function such that mice had constant expectation of when the next trial would begin. The tasks were purely classical conditioning: the behaviour of the mice had no effect on the outcomes. Animals performed between 300 and 700 trials per session.

The dopamine identification experiment (Fig. 1) included three trial types, randomly intermixed. In trial type 1 (45% of all trials), an odour was delivered for 1 s, followed by a 0.5 s delay and a reward chosen pseudo-randomly from the following set: 0.1, 0.3, 1.2, 2.5, 5, 10, or 20 µl. The frequency of each reward size was chosen to make the average reward approximately 5 µl. Reward sizes were determined by the

length of time the water valve remained open: 4, 12, 25, 45, 75, 140, or 250 ms, respectively. In trial type 2 (45% of all trials), rewards of various sizes were delivered without any preceding odour. The reward sizes were identical to trial type 1. In these trials, the reward itself was considered the start of the trial, to ensure a flat hazard function. Comparing trial types 1 and 2 allowed us to determine how a constant level of expectation modulated responses to different sizes of reward. In trial type 3 (10% of all trials), a different odour was delivered, which was followed by no outcome. This trial type was included to ensure that the animals learned the task: they began to lick after the odour in trial type 1 but not after the odour in trial type 3 (Extended Data Fig. 8a).

The GABA stimulation experiment (Fig. 2) mimicked the dopamine identification experiment, but instead of delivering a reward-predicting odour, we used a blue laser to activate VTA GABA neurons directly. The experiment included three randomly interleaved trial types. In trial type 1 (5% of trials), rewards were delivered unexpectedly, in the absence of laser stimulation. Reward sizes were chosen pseudo-randomly from the following set: 0.3, 1.2, 2.5, 5, or 10 µl. Each reward size was equally frequent. In trial type 2 (5% of trials), rewards were also delivered unexpectedly, but now in the presence of laser stimulation. The laser was delivered at 40 Hz for a total of 1 s, and reward was delivered in the middle of this period. In trial type 3 (90% of trials), laser was delivered at 40 Hz for a total of 1 s, but no reward was delivered. The reason for the prevalence of trial type 3 was to ensure that mice did not associate the laser (which they might have seen, despite attempts to mask the light by painting the fibre black) with reward delivery.

In the GABA inhibition experiment (Fig. 3), each trial began with one of two odours, selected pseudo-randomly. One odour predicted water reward with 10% probability and the other odour predicted water reward with 90% probability. In 25% of these trials, 1 s of continuous green laser was administered, beginning at odour offset and lasting until 0.5 s after reward was delivered. This encompassed both the delay between odour and reward (1–1.5 s) and the reward response period (1.5–2 s), which are the times in which VTA GABA neurons normally fire<sup>6</sup>. Laser stimulation did not affect licking behaviour (Extended Data Fig. 8c). At the beginning and end of each recording session, we delivered 1 s periods of green laser without any odours or rewards, to assess how GABA inhibition modulated dopamine baseline activity.

The behavioural experiment (Fig. 4) included four trial types, each associated with a different odour. The four trial types were pseudo-randomly interleaved and equally likely. Odour A was associated with big reward (5 µl), odours B and D were associated with small reward (2 µl), and odour C was associated with no reward. After training, when the mice consistently associated the odours with reward (as demonstrated by their anticipatory licking behaviour), blue laser was paired with odour D trials. Laser was delivered for 2.5 s, beginning 0.5 s after odour onset and ending 0.5 s after reward onset. The intensity of light was modulated in a ramping fashion (see below). After six to eight sessions using the laser, the laser was turned off for the remaining four or five sessions, allowing us to examine whether the effect of laser stimulation would persist even in the absence of laser. Additionally, to clarify whether behaviour changes reflected learning or a direct effect of VTA GABA stimulation on licking, we included probe trials in the final two or three laser sessions. During these probe sessions, 10% of odour B trials randomly received laser stimulation, and 10% of odour D trials randomly omitted the laser.

**Electrophysiology.** Recording techniques were based on a previous study<sup>6</sup>. Briefly, we recorded extracellularly from VTA using a custom-built, screw-driven microdrive containing six or eight tetrodes (Sandvik) glued to a 200 µm optic fibre (Thorlabs). Tetrodes were affixed to the fibre so that their tips extended 300–600 µm from the end of the fibre. Neural and behavioural signals were recorded with a DigiLynx recording system (Neuralynx) or a custom-built system using a multi-channel amplifier chip (RHA2116, Intan Technologies) and data acquisition device (PCIe-6351, National Instruments). Broadband signals from each wire were filtered between 0.1 and 9,000 Hz and recorded continuously at 32 kHz. To extract spike timing, signals were band-pass-filtered between 300 and 6,000 Hz and sorted offline using SpikeSort3D (Neuralynx) or MClust-3.5 (A. D. Redish). At the end of each session, the fibre and tetrodes were lowered by 40–80 µm to record new units the next day.

To be included in the data set, a neuron had to be well isolated (*L*-ratio (ref. 36) < 0.05) and recorded within 0.5 mm of a light-identified or putative dopamine neuron, to ensure that it was recorded in VTA. Recording sites were also verified histologically with electrolytic lesions using 10–15 s of 30 µA direct current.

**Laser delivery.** To identify neurons as dopaminergic or GABAergic, we used ChR2 to observe laser-triggered spikes<sup>6,37,38</sup>. The optical fibre was coupled with a diode-pumped solid-state laser with analogue amplitude modulation (Laserglow Technologies). At the beginning and end of each recording session, we delivered trains of ten blue (473 nm) light pulses, each 5 ms long, at 1, 10, 20 and 50 Hz, with an intensity of 5–20 mW mm<sup>-2</sup> at the tip of the fibre. Spike shape was measured using a broadband signal (0.1–9,000 Hz) sampled at 32 kHz.



In the GABA stimulation experiment (Fig. 2), we used the same blue laser to deliver 40 pulses (duration 5 ms, 40 Hz) during selected trials. In the GABA inhibition experiment (Fig. 3), we used one of two methods of laser delivery. For seven mice (Extended Data Fig. 9a–c), we used an electronic shutter (Vincent Associates) to deliver 1 s intervals of continuous green laser (532 nm, Laserglow Technologies), with an intensity of  $\sim 50 \text{ mW mm}^{-2}$  at the tip of the fibre. For a separate group of two mice (Extended Data Figs 9d–f and 10), we instead modulated laser intensity in an analogue fashion, beginning at zero intensity 0.5 s after odour onset, smoothly increasing intensity to a peak of  $50 \text{ mW mm}^{-2}$  at reward delivery, and then gradually decreasing over the next 0.5 s. This ramping protocol was also used for the behavioural experiment (Fig. 4), with a 473 nm laser (OptoEngine) and beam splitter (Doric Lenses) to deliver blue light bilaterally. The ramping intensity profile was chosen to approximate the response pattern of VTA GABA neurons<sup>6</sup>.

**Data analysis.** Peristimulus time histograms were constructed using 1 ms bins and then convolved with a function resembling a postsynaptic potential,  $(1 - \exp(-t)) \times (\exp(-t/20))$ , for time  $t$  in milliseconds. Average firing rates in response to reward were calculated using a 600 ms window after reward onset for the dopamine identification and GABA stimulation experiments, and a 500 ms window after reward onset for the GABA inhibition experiment. These windows were chosen to reflect the full duration of the neural response to reward. Window sizes ranging from 300 to 1,000 ms were attempted and gave qualitatively similar results. To calculate reward response, we subtracted baseline firing (averaged over 1 s before trial onset). Calculating the baseline using different windows (for example, 600 ms before reward onset) did not change the results. To ensure reliability, analyses of particular trial types only included neurons that were recorded during at least five presentations of that trial type.

To identify neurons as dopaminergic or GABAergic, we used the stimulus-associated spike latency test<sup>38</sup> to determine whether light pulses significantly changed a neuron's spike timing (Extended Data Fig. 3). We used a significance value of  $P < 0.001$ . To ensure that spike sorting was not contaminated by light artefacts, we also calculated waveform correlations between spontaneous and light-evoked spikes, as described<sup>6</sup>. All light-identified neurons had Pearson's correlation coefficients  $> 0.9$ .

In all three recording experiments, we identified putative dopamine and GABA neurons on the basis of their firing patterns through an unsupervised clustering approach (Extended Data Figs 2 and 9), similar to a previous study<sup>6</sup>. Briefly, receiver-operating characteristic (ROC) curves for each neuron were calculated by comparing the distribution of firing rates across trials in 100 ms bins (starting 1 s before expected reward and ending 1 s after expected reward) to the distribution of baseline firing rates (1 s before trial onset). Principal component analysis was calculated using the singular value decomposition of the area under the ROC. Hierarchical clustering was then done using the first three principal components of the area under the ROC using a Euclidean distance metric and complete agglomeration method.

As described<sup>6</sup>, this method produced three clusters: one with phasic excitation to reward (type 1), one with sustained excitation to reward expectation (type 2), and one with sustained suppression to reward expectation (type 3). Type 1 neurons were classified as putatively dopaminergic. Forty out of 43 light-identified dopamine neurons fell into this cluster; the other three light-identified dopamine neurons showed phasic suppression to reward and were clustered as type 3. Since these three dopamine neurons showed qualitatively different responses than the others, they were not included in the data set. Note that although we focus on identified dopamine neurons, our main findings are identical if we include all putative dopamine neurons (Extended Data Fig. 6b, c).

Type 2 neurons were classified as putatively GABAergic. Of 14 identified GABA neurons, 11 were clustered as type 2; the other three were inhibited by reward and were clustered as type 3. Again, these three GABA neurons were not included in the data set. Unlike type 1 neurons, type 2 neurons did not respond to either expected or unexpected reward in a consistently size-dependent fashion (Extended Data Fig. 5). This contrasts with their delay activity, which increases with increasing reward expectation<sup>6</sup>.

The distribution of neurons across mice for all recording experiments is provided in Supplementary Table 1.

To determine the dose-response of dopamine neurons and see whether expectation caused a subtractive or divisive effect (Fig. 1c), we based our analysis on a previous study<sup>39</sup>. We first fitted a hyperbolic ratio function (Hill function) to the unexpected reward data:

$$f(r) = f_{\max} \left( \frac{r^{0.5}}{r^{0.5} + \sigma^{0.5}} \right) \quad (1)$$

The function had two free parameters:  $f_{\max}$ , the saturating firing rate; and  $\sigma$ , the reward size that elicits half-maximum firing rate. We chose an exponent of 0.5

after fitting the data with exponents ranging from 0.1 to 2.0 (in steps of 0.1), and finding the exponent with the lowest mean squared error. Note that the Hill function is not the only possible function that could fit our data. For example, the power function  $f(r) = ar^k$ , where  $a = 3.73$  and  $k = 0.39$ , was also excellent. However, this function does not saturate, so we thought it was less likely to represent neuronal responses. Our conclusions do not depend on the exact function chosen to fit the data.

After fitting the unexpected reward data, we explored what simple transformation could best mimic the effect of expectation. We tested four options: input subtraction, input division, output subtraction, and output division (Extended Data Fig. 6a). Specifically, we evaluated the following four models<sup>39</sup>:

$$\text{Input subtraction: } f(r) = f_{\max} \left( \frac{(r-E)^{0.5}}{(r-E)^{0.5} + \sigma^{0.5}} \right) \quad (2)$$

$$\text{Input division: } f(r) = f_{\max} \left( \frac{r^{0.5}}{r^{0.5} + \sigma^{0.5} + E^{0.5}} \right) \quad (3)$$

$$\text{Output subtraction: } f(r) = f_{\max} \left( \frac{r^{0.5}}{r^{0.5} + \sigma^{0.5}} \right) - E \quad (4)$$

$$\text{Output division: } f(r) = \left( \frac{1}{E^{0.5} + 1} \right) \times f_{\max} \left( \frac{r^{0.5}}{r^{0.5} + \sigma^{0.5}} \right) \quad (5)$$

In each case, we used the  $f_{\max}$  and  $\sigma$  values determined by the unexpected reward data. The only new parameter was the expectation factor  $E$ , which we fitted separately for each of the four models. Output subtraction consistently gave the best fit (lowest mean squared error), for the population and most individual neurons. The next best model was generally output division. We statistically compared model-fits using a bootstrapping analysis: we resampled the data 1,000 times and determined for each resample the mean squared error for both output subtraction and output division. We calculated the  $P$  value by counting the number of resamples when the mean squared error was better for output division than for output subtraction (for example, if 1 resample out of 1,000 preferred output division over output subtraction,  $P = 0.001$ ; Extended Data Fig. 6c). These steps were repeated for putative dopamine neurons in the GABA stimulation experiment (Fig. 2d).

As a complementary analysis to determine whether expectation had a subtractive or divisive effect on dopamine reward responses, we calculated the difference between unexpected and expected reward responses for different reward sizes (Fig. 1d). We then ran a linear regression to determine whether the slope of this difference was significantly different from zero. A slope of zero would be consistent with output subtraction, as expectation would have the same effect on all responses. A slope greater than zero would be consistent with output division, as expectation would have a larger effect on larger responses. All but five of the light-identified dopamine neurons had a slope no different from zero (Extended Data Fig. 4c).

In our GABA stimulation and inhibition experiments, we wanted to ensure that laser delivery affected phasic dopamine responses in addition to shifting baseline dopamine activity. First, we identified putative dopamine neurons that did not significantly change their baseline firing upon laser delivery. To do so, we compared firing rates in the 0.5 s before reward delivery on laser trials versus no-laser trials. Neurons with  $P > 0.05$  (Wilcoxon rank-sum) were identified as unaffected by laser delivery. In both the GABA stimulation and GABA inhibition experiments, these neurons continued to be affected at the time of reward (Extended Data Fig. 7a, e). Second, we recorded from putative dopamine neurons while manipulating VTA GABA activity outside the task (Extended Data Fig. 7c, g). This gave us an unbiased sense of how VTA GABA stimulation or inhibition affected dopamine baseline responses. We then subtracted these laser-alone trials from trials where laser was delivered during reward (Extended Data Fig. 7b, f). Any remaining change at the time of reward should not be due to a baseline shift.

Interestingly, the baseline shift may have been an artefact of the type of laser stimulation we applied. In a separate experiment ( $n = 2$  mice, Extended Data Fig. 9d–f), we applied the laser so that light intensity would ramp up rather than remain constant over the course of a trial, more closely mimicking the physiological responses of VTA GABA neurons. We found that this ramping stimulation successfully inhibited putative VTA GABA neurons ( $P = 0.001$ ,  $t$ -test, Extended Data Fig. 10a, b) and increased reward responses in putative dopamine neurons ( $P < 0.001$ ,  $t$ -test, Extended Data Fig. 10c, d) without causing a baseline shift.

To assess how well VTA GABA stimulation mimics odour expectation, we also directly compared the magnitude of change in dopamine responses in both experiments. In the odour-based experiment (Fig. 1), the average suppression of dopamine reward responses was 52.7%, compared with 43.5% for the VTA GABA stimulation experiment (Fig. 2;  $P < 0.05$ ,  $t$ -test). This difference may be accounted for by variation among putative dopamine neurons in their response to laser. Although 40 out of 45 putative dopamine neurons were suppressed by GABA stimulation, 5 were activated, perhaps through disynaptic disinhibition, as VTA GABA neurons are known to synapse onto each other as well as onto dopamine neurons<sup>13</sup>. In addition, there may be other neurons, besides VTA GABA neurons, that help suppress dopamine responses when reward is expected.

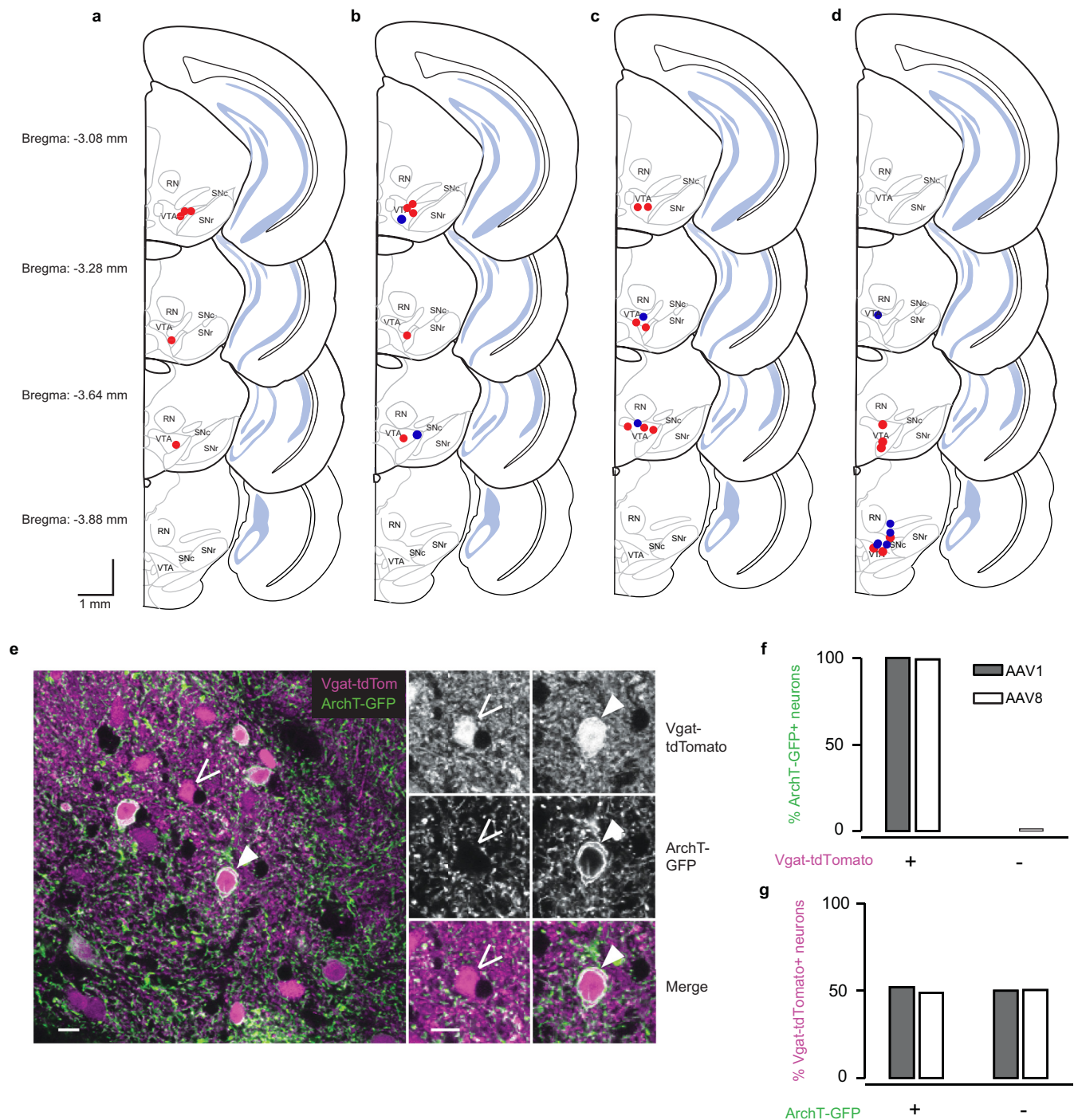
Although we focus on changes in the magnitude of dopamine neuron responses, this was not the only effect of reward expectation. Notably, the latency to peak response was also extended, from an average of 67.6 ms to 95.4 ms ( $P = 0.001$ ,  $t$ -test). The latency increased in 37 out of 40 dopamine neurons that we recorded. The downstream consequences of this change in latency remain to be elucidated.

Comparisons were performed with  $t$ -tests (for population data) or Wilcoxon rank-sum tests (for individual neuron data), with corrections for multiple comparisons (Bonferroni or Tukey). Correlations were done with Pearson's rho.  $P$  values less than 0.05 were considered significant, unless otherwise noted. Given pilot data showing effects of optogenetic manipulation of about 2 spikes per second, with variability of about 3 spikes per second, 36 neurons were required for 80% power to detect the effect. Given about ten neurons of each type per mouse, we aimed for at least four mice per experiment. Analyses were done with Matlab (Mathworks).

In the behavioural experiment (Fig. 4), the strength of the learned association between each odour and reward was estimated by counting the number of anticipatory licks over the 2 s from odour onset to reward delivery. For the analysis in Fig. 4b and Extended Data Fig. 8d, e, we excluded data from probe trials. Population results were examined using a mixed-effects linear model. The fixed effects included trial type and a binary variable indicating whether the session included laser delivery. The random effect was mouse identity. The outcome of interest was an interaction between trial type and laser. Results were robust to different choices of window for counting anticipatory licks.

**Immunohistochemistry.** After recording for 4–8 weeks, mice were given an overdose of ketamine/medetomidine, exsanguinated with saline, and perfused with 4% paraformaldehyde. Brains were cut in 100  $\mu$ m coronal sections on a vibrotome and immunostained with antibodies to tyrosine hydroxylase (AB152, 1:1,000, Millipore) to visualize dopamine neurons and 49,6-diamidino-2-phenylindole (DAPI, Vectashield) to visualize nuclei. Virus expression was determined through eYFP fluorescence. Slides were examined to verify that the optic fibre track was among VTA dopamine neurons and in a region expressing the virus. For the GABA inhibition experiment, two *Vgat*-tdTomato mice were injected with AAV-FLEX-ArchT-GFP to determine the selectivity and efficiency of ArchT expression in VTA GABA neurons (Extended Data Fig. 1e–g). One mouse was injected with AAV serotype 1 and the other with AAV serotype 8. For the figure, brightness and contrast were adjusted in Adobe Photoshop.

31. Bäckman, C. M. *et al.* Characterization of a mouse strain expressing Cre recombinase from the 3' untranslated region of the dopamine transporter locus. *Genesis* **44**, 383–390 (2006).
32. Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71**, 142–154 (2011).
33. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
34. Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28**, 7025–7030 (2008).
35. Uchida, N. & Mainen, Z. F. Speed and accuracy of olfactory discrimination in the rat. *Nature Neurosci.* **6**, 1224–1229 (2003).
36. Schmitzer-Torbert, N. & Redish, A. D. Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. *J. Neurophysiol.* **91**, 2259–2272 (2004).
37. Lima, S. Q., Hromádka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLoS One* **4**, e6099 (2009).
38. Kvitsiani, D. *et al.* Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).
39. Olsen, S. R., Bhandawat, V. & Wilson, R. I. Divisive normalization in olfactory population codes. *Neuron* **66**, 287–299 (2010).

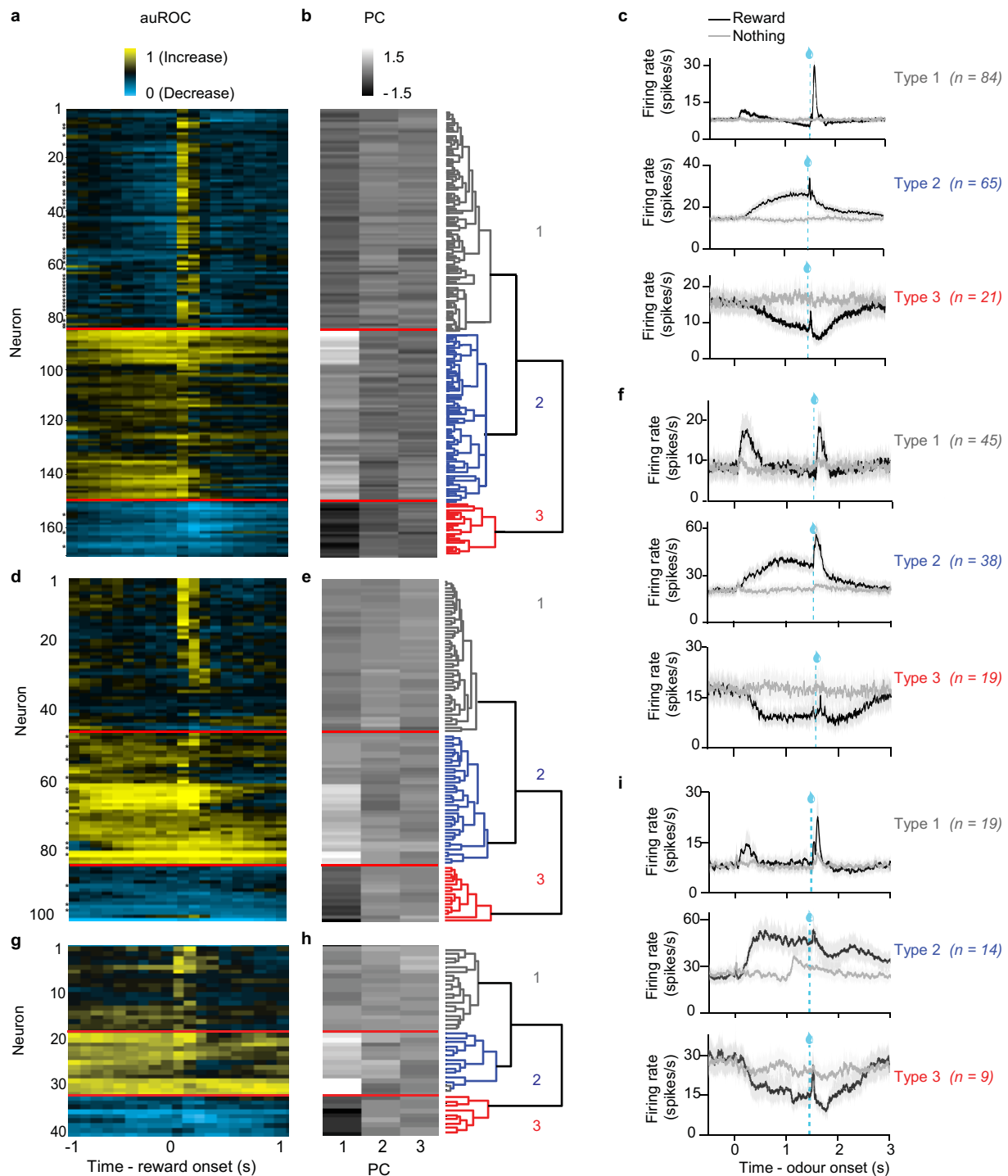


#### Extended Data Figure 1 | Recording sites and ArchT expression.

**a–d**, Schematic of recording locations for mice used in the dopamine identification task ( $a$ ,  $n = 5$ ), the GABA stimulation task ( $b$ ,  $n = 7$ ), the GABA inhibition task ( $c$ ,  $n = 9$ ), and the behavioural task ( $d$ ,  $n = 12$ ). Red, experimental mice expressing ChR2 in VTA GABA neurons ( $n = 5$ ). Blue, control mice expressing GFP in VTA GABA neurons ( $n = 2$ ). **b**, Red, experimental mice expressing ChR2 in VTA GABA neurons ( $n = 5$ ). Blue, control mice expressing GFP in VTA GABA neurons ( $n = 2$ ). **c**, Red, mice in which laser was delivered at continuous intensity ( $n = 7$ ). Blue, mice in which laser was delivered with ramping intensity ( $n = 2$ ). **d**, Red, experimental mice expressing ChR2 in VTA GABA neurons ( $n = 6$ ). Blue, control mice expressing GFP in VTA GABA neurons ( $n = 6$ ). **e–g**, Selectivity and efficiency of ArchT expression. **e**, Representative merged image (one of 30 z-stacks).

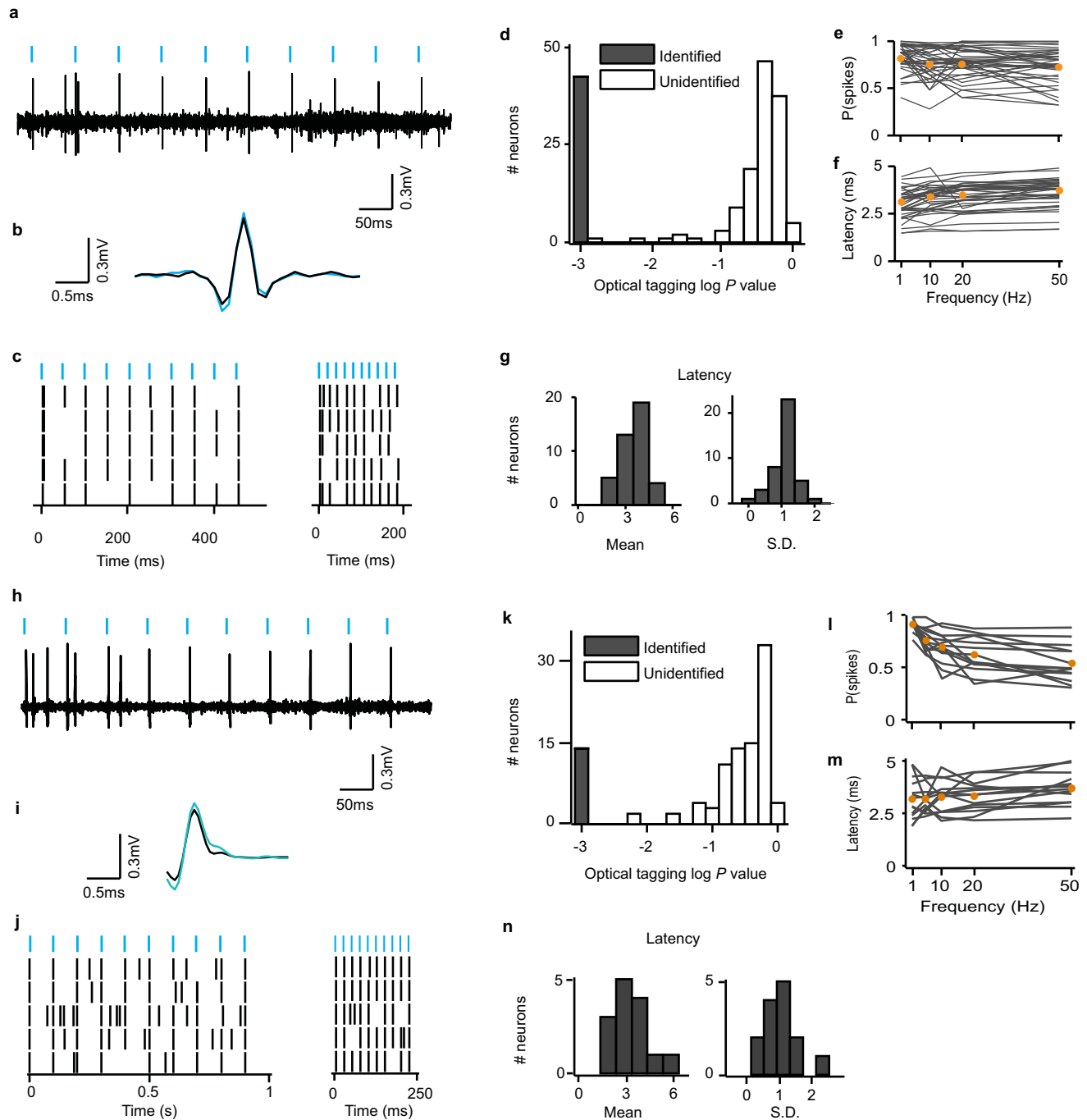
Magenta, *Vgat-tdTomato*; green, ArchT-GFP. Open arrow, neuron expressing *Vgat-tdTomato* but not ArchT-GFP. Closed arrow, neuron expressing both *Vgat-tdTomato* and ArchT-GFP. Scale bar, 10  $\mu$ m. **f**, Selectivity of infection to GABA neurons: percentage of ArchT-GFP-expressing neurons ( $n = 131$  neurons for AAV1 and 165 neurons for AAV8) that were positive for *Vgat-tdTomato*. Filled bars, *Vgat-tdTomato* mouse injected with AAV1-FLEX-ArchT-GFP. Empty bars, *Vgat-tdTomato* mouse injected with AAV8-FLEX-ArchT-GFP. **g**, Efficiency of infection: percentage of *Vgat-tdTomato*-expressing neurons ( $n = 278$  neurons for AAV1 and 283 neurons for AAV8) that were positive for ArchT-GFP.





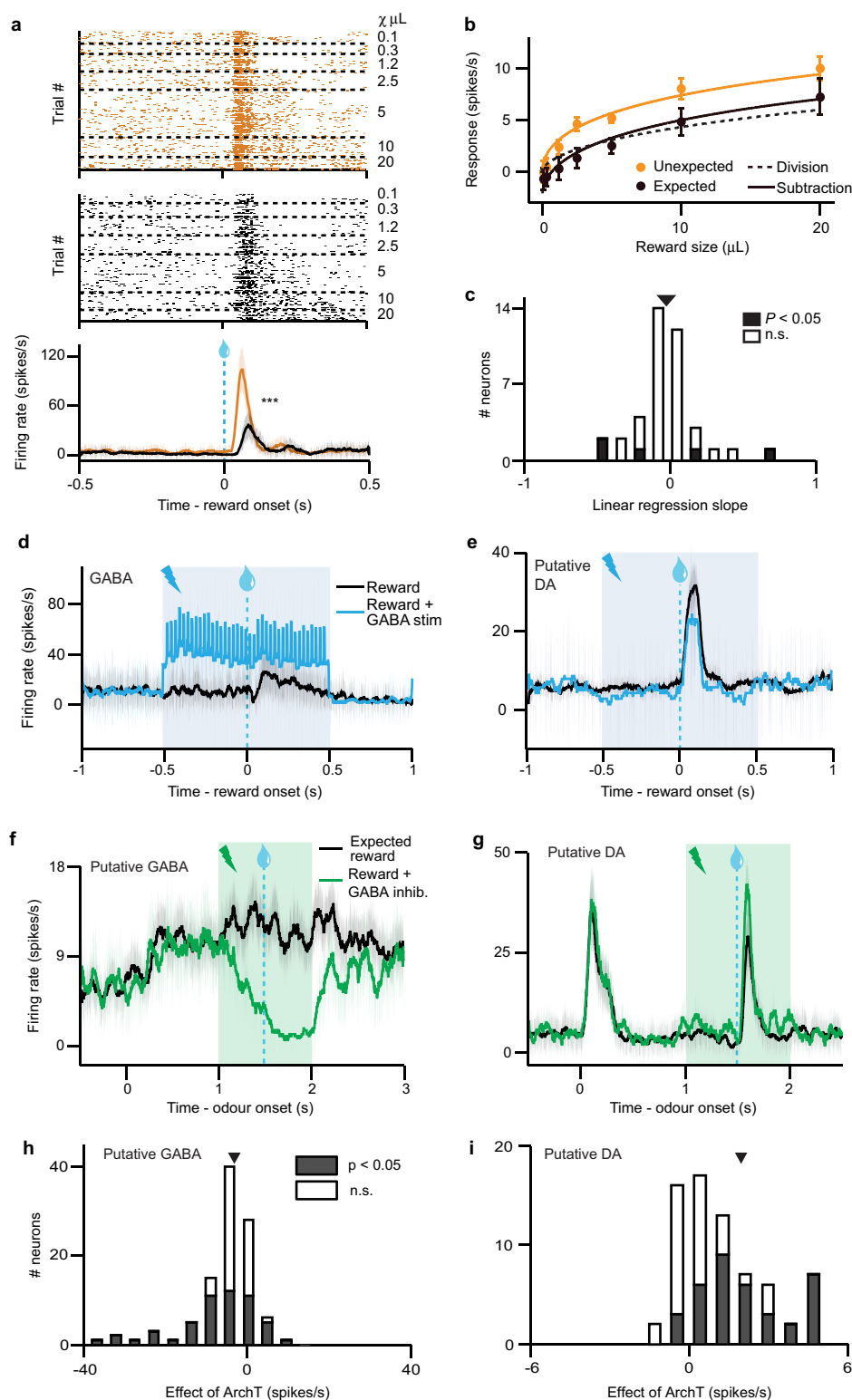
**Extended Data Figure 2 | Neuron classification for dopamine identification and GABA stimulation experiments.** **a–c**, Dopamine identification experiment. **d–f**, ChR2-expressing animals in GABA stimulation experiment. **g–i**, GFP-expressing control animals in GABA stimulation experiment. **a, d, g**, Responses of all VTA neurons recorded in the tasks. Each row reflects the area under the ROC values for a single neuron in the second before and after delivery of expected reward. Baseline is taken as 1 s before odour onset.

Yellow, increase from baseline; cyan, decrease from baseline. Light-identified neurons are denoted by an asterisk to the left of each column. **b, e, h**, The first three principal components of the area under the ROC curves. These values were used for unsupervised hierarchical clustering, as shown in the dendrogram on the right. **c, f, i**, Average firing rates for the three clusters of neurons in each task. Odour was delivered for 1 s, followed by a 0.5 s delay and then reward delivery.



**Extended Data Figure 3 | Light identification of dopamine and GABA neurons.** **a**, Raw signal from one example light-identified dopamine neuron. Blue bars, light pulses. **b**, For the same neuron, mean waveforms for spontaneous (black) and light-evoked (blue) action potentials. **c**, For the same neuron, raster plots for 20 Hz (left) and 50 Hz (right) laser stimulation. Each row is one trial of laser stimulation. **d**, Histogram of log  $P$  values for each neuron recorded in the dopamine identification experiment ( $n = 170$ ). The  $P$  values were derived from the stimulus-associated spike latency test (see Methods).

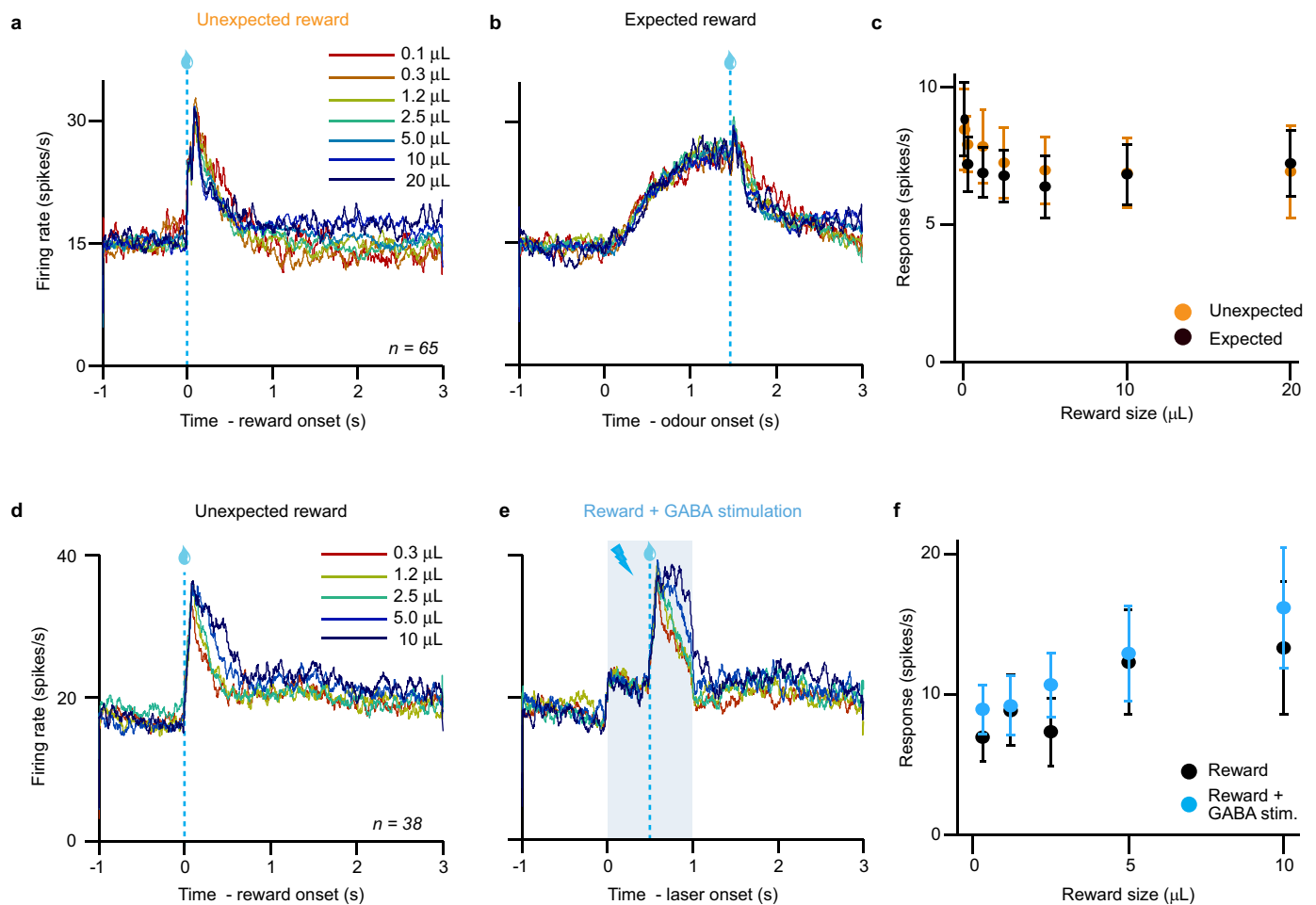
Neurons with  $P < 0.001$  and waveform correlations  $> 0.9$  were considered identified (filled bars). **e**, **f**, For light-identified neurons, probability of spiking (**e**) and latency to first spike (**f**) after laser pulses at different frequencies. Orange circles, mean across neurons. **g**, Histogram of mean latencies (left) and latency standard deviations (right) in response to laser stimulation for all light-identified dopamine neurons in the variable-reward task. **h–n**, Same conventions as **a–g**, but for neurons recorded in the GABA stimulation task ( $n = 102$ ).



**Extended Data Figure 4 | Individual neuron analysis from all recording experiments.** **a–c**, Results from dopamine identification experiment (Fig. 1). **d, e**, Results from GABA stimulation experiment (Fig. 2). **f–i**, Results from GABA inhibition experiment (Fig. 3). **a**, Raster plots (top and middle) and firing rate (bottom) of representative dopamine neuron in response to unexpected (orange) or temporally expected (black) reward. \*\*\* $P < 0.001$ ,  $t$ -test. **b**, For the same neuron, responses (mean  $\pm$  s.e.m. across trials) to each reward size. Orange line, fit for unexpected reward. Dotted black line, divisive transformation. Solid black line, subtractive transformation. **c**, Individual

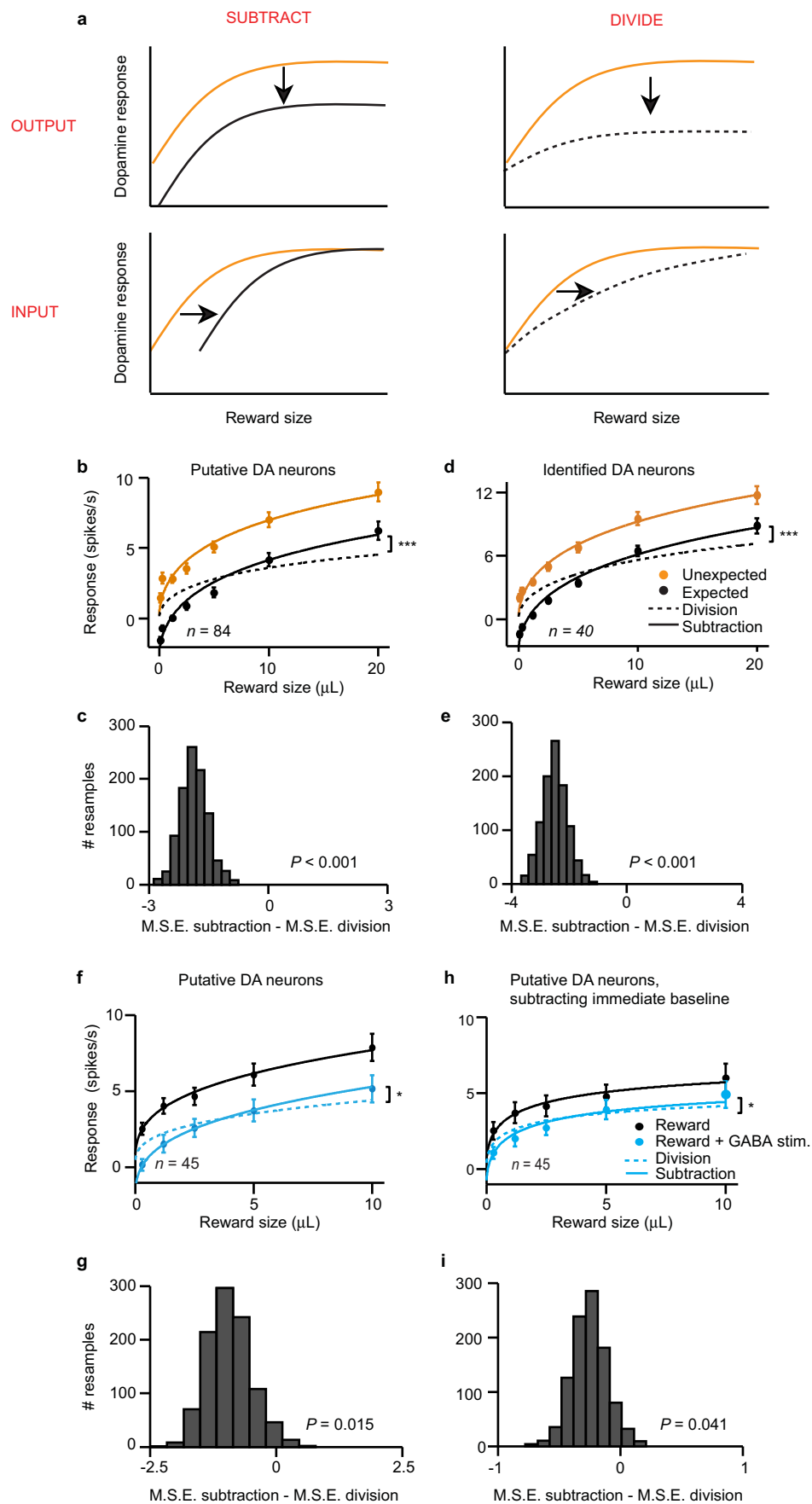
neuron regression slopes for the analysis in Fig. 1d. Empty bars, slope not different from zero ( $P > 0.05$ ). Filled bars,  $P < 0.05$ . Triangle, mean slope. **d, e**, Firing rate of representative VTA GABA (**d**) and putative dopamine (**e**) neuron with (blue) and without (black) ChR2 stimulation. Light blue box, laser delivery. **f, g**, Firing rate of representative VTA GABA (**f**) and putative dopamine (**g**) neuron during odour B trials with (green) or without (black) laser delivery. **h, i**, Histogram of putative GABA (**h**) and dopamine (**i**) neuron responses to laser delivery. Filled bars, significant effect of laser ( $P < 0.05$ , Wilcoxon rank-sum); empty bars,  $P > 0.05$ . Triangle, mean.





**Extended Data Figure 5 | VTA GABA activity does not vary consistently with reward size.** **a–c**, Putative GABA neurons in the dopamine identification experiment (Fig. 1). **d–f**, Putative GABA neurons in the GABA stimulation experiment (Fig. 2). **a, b**, Average firing rate of putative GABA neurons to unexpected (**a**) or temporally expected (**b**) rewards of various sizes. **c**, Population responses (mean  $\pm$  s.e.m. across putative GABA neurons) for different reward sizes. Orange, unexpected reward. Black, temporally expected

reward. Responses were averaged over a 600 ms window after reward delivery. **d, e**, Average firing rate of putative GABA neurons to rewards of various sizes, delivered with (**e**) or without (**d**) optogenetic GABA stimulation. **f**, Population responses (mean  $\pm$  s.e.m. across putative GABA neurons) for different reward sizes. Blue, reward with laser stimulation. Black, reward without laser stimulation. Responses were averaged over a 600 ms window after reward delivery.

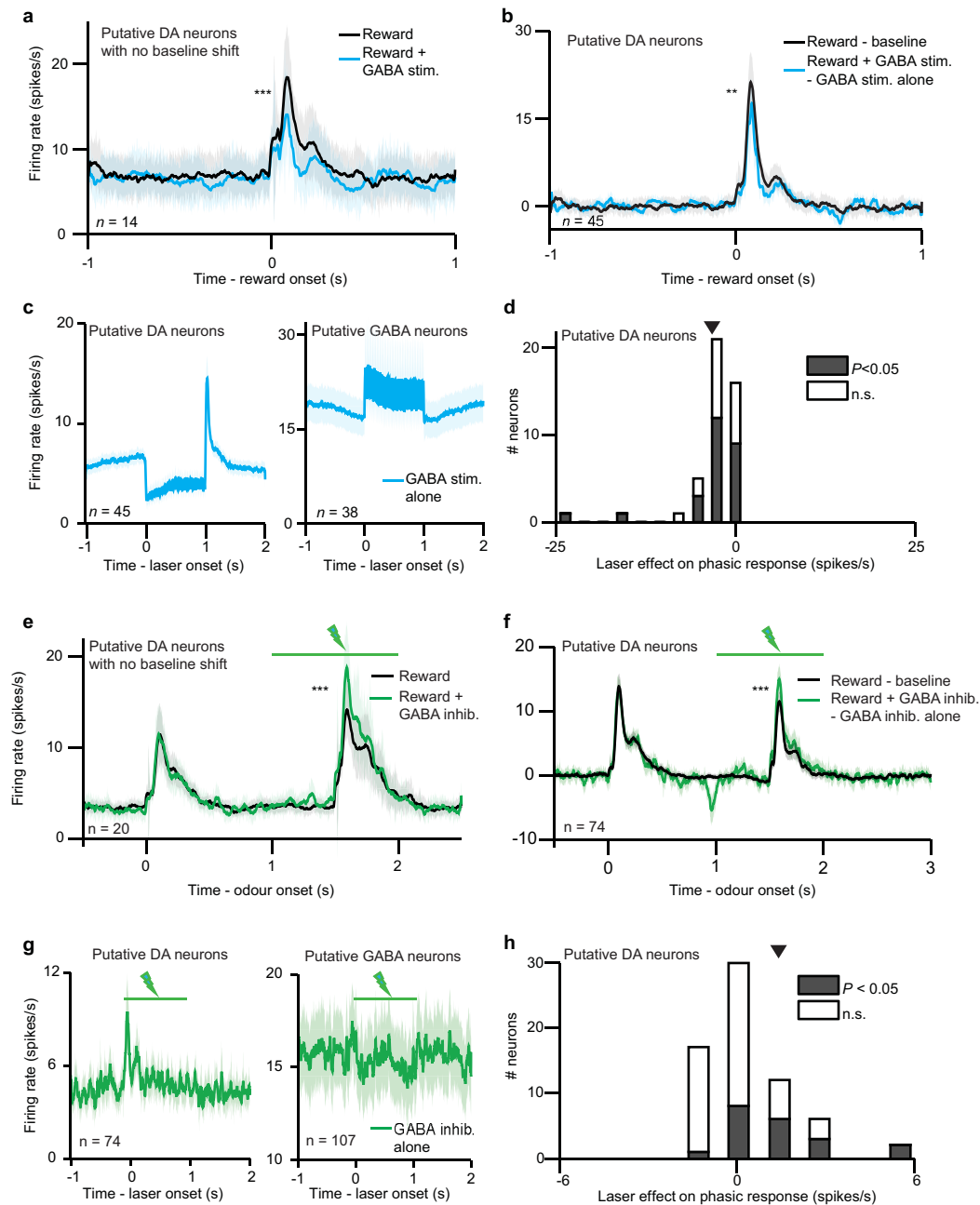


**Extended Data Figure 6 | Statistical test for subtraction versus division.**

**a**, To understand how dopamine neurons compute reward prediction error, we first determined how dopamine neurons respond to various sizes of unexpected reward (schematized as orange curves). We then taught the mice to expect reward and observed how expectation shifted this dose–response (black curves). We modelled four types of shift: output subtraction (top left), input subtraction (bottom left), output division (top right), and input division (bottom right). Output subtraction was consistently the best fit. For equations, see Methods. Analysis adapted from a previous study<sup>39</sup>. **b–e**, Results from dopamine identification experiment. **f–i**, Results from GABA stimulation experiment. **b, c**, Results from all putative dopamine neurons ( $n = 84$ ). \*\*\* $P < 0.001$ , bootstrap. **d, e**, Results from light-identified dopamine neurons

( $n = 40$ ). \*\*\* $P < 0.001$ , bootstrap. **f, g**, Results from putative dopamine neurons in the GABA stimulation experiment ( $n = 45$ ). \* $P < 0.05$ , bootstrap. **h, i**, Results from putative dopamine neurons in the GABA stimulation experiment, subtracting the 500 ms period immediately before reward delivery. This takes into account the laser-induced baseline shift in dopamine responses. \* $P < 0.05$ , bootstrap. **b, d, f, h**, Average responses (mean  $\pm$  s.e.m. across neurons) to different sizes of reward, with fits for output subtraction (solid line) and output division (dotted line). **c, e, g, i**, Results of bootstrapping analysis. For each resample, we compared the mean squared error for the subtractive fit with the mean squared error for the divisive fit. Negative numbers favour subtraction.  $P$  values were calculated as the proportion of resamples in which division was a better fit than subtraction.

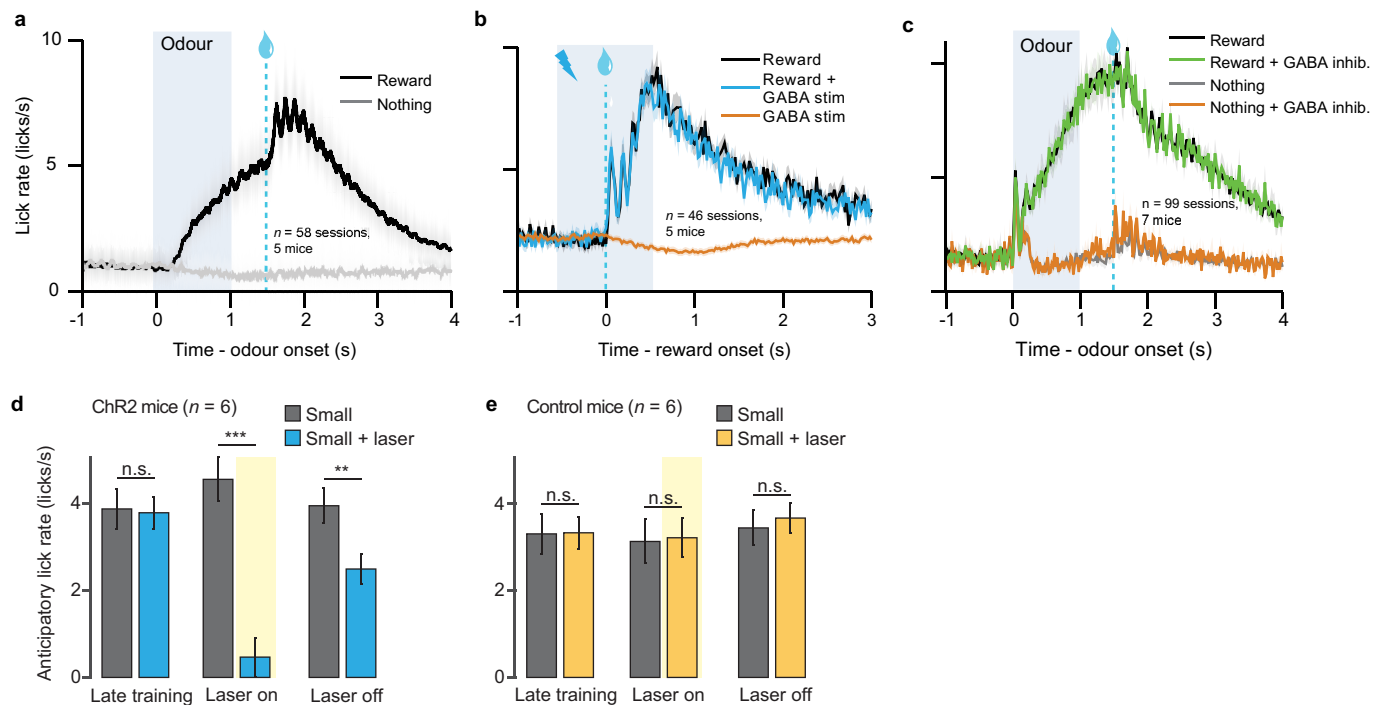




### Extended Data Figure 7 | Laser effect is more than a baseline shift.

**a–d**, Results from GABA stimulation experiment. **e–h**, Results from GABA inhibition experiment. **a**, Firing rate (mean  $\pm$  s.e.m.) of putative dopamine neurons that did not show a significant baseline shift. \*\*\* $P < 0.001$ ,  $t$ -test. **b**, To visualize whether GABA stimulation preferentially affected phasic dopamine responses in addition to baseline firing rates, we took the activity in Fig. 2c and subtracted the trials when laser was delivered alone. Any remaining change at the time of reward could not be due to a baseline shift. \*\* $P = 0.01$ ,  $t$ -test. **c**, Firing rate (mean  $\pm$  s.e.m.) of putative dopamine (left) and GABA (right)

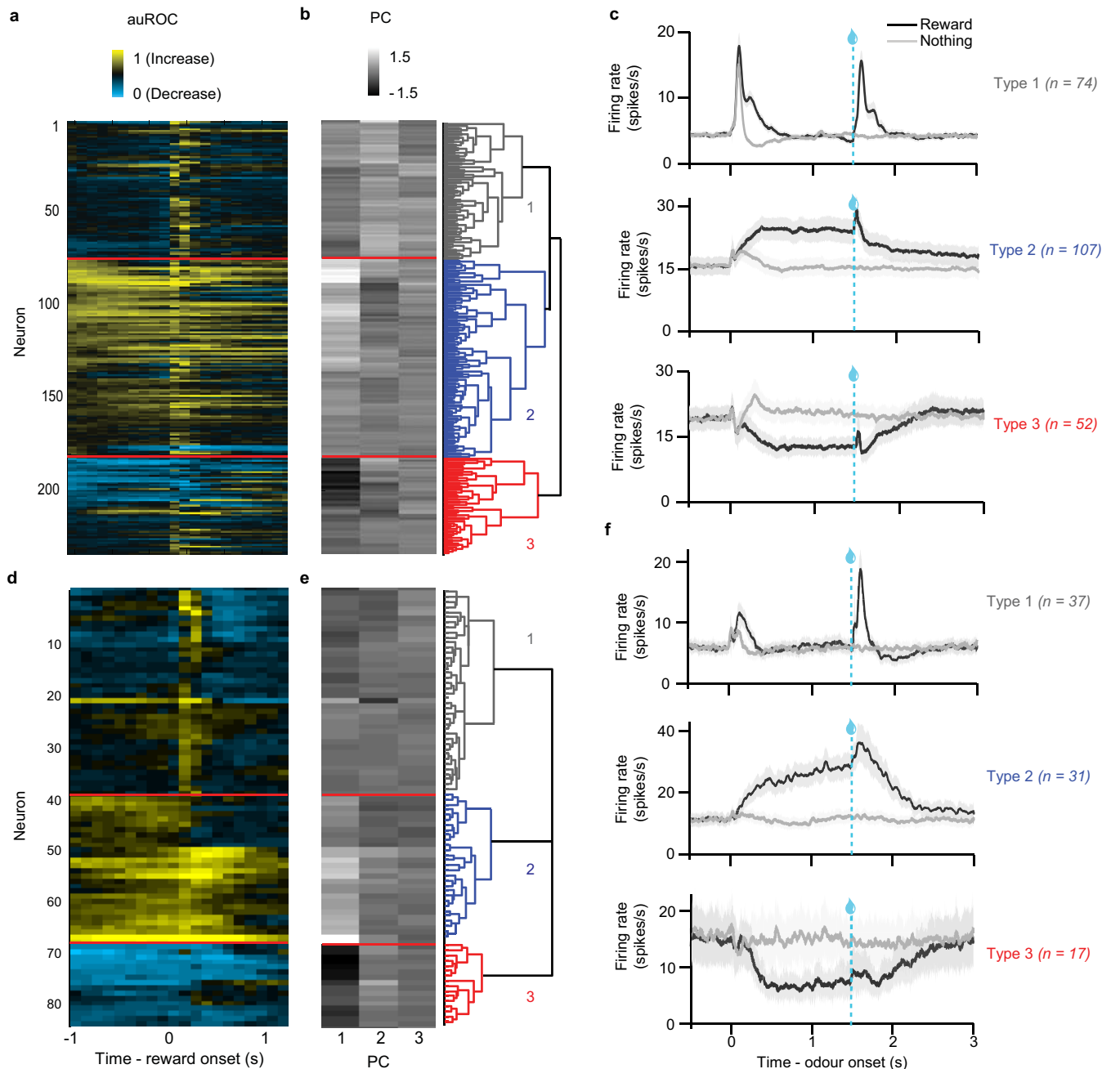
neurons in trials where laser was delivered in the absence of reward. This dopamine response was subtracted to calculate the firing rates in **b**. **d**, Histogram of the phasic effect of GABA stimulation. The values were calculated by subtracting the black line from the blue line in **b**. Empty bars, slope not different from zero ( $P > 0.05$ , Wilcoxon rank-sum test). Filled bars, slope different from zero ( $P < 0.05$ ). Triangle, mean ( $P < 0.001$ ,  $t$ -test). **e–h**, Same conventions as **a–d**, but for the GABA inhibition experiment. \*\*\* $P < 0.001$ ,  $t$ -test.



#### Extended Data Figure 8 | Behavioural performance on all four experiments.

**a**, In the dopamine identification task (Fig. 1), lick rates (mean  $\pm$  s.e.m. across sessions) for odours predicting reward (black) or nothing (grey). **b**, In the GABA stimulation task (Fig. 2), lick rates (mean  $\pm$  s.e.m. across sessions) for reward alone (black), reward + GABA stimulation (blue), or GABA stimulation alone (orange). **c**, In the GABA inhibition task (Fig. 3), lick rates (mean  $\pm$  s.e.m. across sessions) for the odours predicting reward with 90% probability (black) and 10% probability (grey). Green laser was delivered to

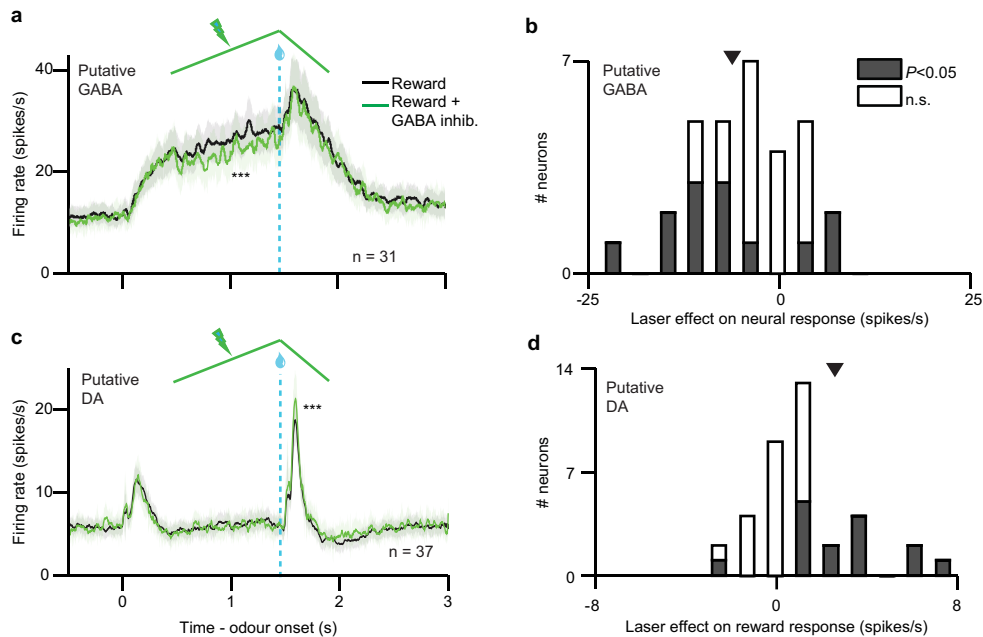
inhibit VTA GABA neurons in 25% of reward (green) and nothing (orange) trials. **d**, **e**, In the bilateral stimulation experiment (Fig. 4), anticipatory licks (mean  $\pm$  s.e.m. across mice) for mice injected with ChR2 (**d**) and GFP (**e**). Grey bars, odour B; blue or orange bars, odour D. Left, last three training sessions before odour D was paired with laser; middle, last three sessions with laser delivery (excluding probe trials); right, last three sessions after laser was turned off. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; paired  $t$ -test.



**Extended Data Figure 9 | Neuron classification for GABA inhibition experiment.** **a–c**, Mice in which laser was delivered with continuous intensity. **d–f**, Mice in which laser was delivered with ramping intensity. **a, d**, Responses of all VTA neurons recorded in the tasks. Each row reflects the area under the ROC values for a single neuron in the second before and after delivery of expected reward. Baseline is taken as one second before odour onset. Yellow,

increase from baseline; cyan, decrease from baseline. **b, e**, The first three principal components of the area under the ROC curves. These values were used for unsupervised hierarchical clustering, as shown in the dendrogram on the right. **c, f**, Average firing rates for the three clusters of neurons in each task. Odour was delivered for 1 s, followed by a 0.5 s delay and then reward delivery.





**Extended Data Figure 10 | Ramping laser stimulation eliminates baseline shift.** **a**, Firing rate (mean  $\pm$  s.e.m.) of putative VTA GABA neurons during odour B trials with (green) or without (black) ramping laser delivery. \*\*\* $P < 0.001$ ,  $t$ -test. **b**, Histogram of putative GABA neuron responses to laser delivery. Responses were averaged over the entire duration of the laser. Filled bars, significant effect of laser ( $P < 0.05$ , Wilcoxon rank-sum test); empty

bars,  $P > 0.05$ . Triangle, mean ( $P < 0.001$ ,  $t$ -test). **c**, Firing rate (mean  $\pm$  s.e.m.) of putative dopamine neurons with (green) or without (black) ramping GABA inhibition. \*\*\* $P < 0.001$ ,  $t$ -test. **d**, Histogram of putative dopamine neuron responses to laser delivery. Responses were averaged over the 0.5 s window after reward delivery. Filled bars, significant effect of laser ( $P < 0.05$ , Wilcoxon rank-sum test); empty bars,  $P > 0.05$ . Triangle, mean ( $P < 0.001$ ,  $t$ -test).

# Evidence for human transmission of amyloid- $\beta$ pathology and cerebral amyloid angiopathy

Zane Jaunmuktane<sup>1</sup>, Simon Mead<sup>2,3,4</sup>, Matthew Ellis<sup>3</sup>, Jonathan D. F. Wadsworth<sup>2,3</sup>, Andrew J. Nicoll<sup>2,3</sup>, Joanna Kenny<sup>2,4</sup>, Francesca Launchbury<sup>3</sup>, Jacqueline Linehan<sup>2</sup>, Angela Richard-Loendt<sup>3</sup>, A. Sarah Walker<sup>5</sup>, Peter Rudge<sup>2,4</sup>, John Collinge<sup>2,3,4</sup> & Sebastian Brandner<sup>1,2,3</sup>

More than two hundred individuals developed Creutzfeldt–Jakob disease (CJD) worldwide as a result of treatment, typically in childhood, with human cadaveric pituitary-derived growth hormone contaminated with prions<sup>1,2</sup>. Although such treatment ceased in 1985, iatrogenic CJD (iCJD) continues to emerge because of the prolonged incubation periods seen in human prion infections. Unexpectedly, in an autopsy study of eight individuals with iCJD, aged 36–51 years, in four we found moderate to severe grey matter and vascular amyloid- $\beta$  (A $\beta$ ) pathology. The A $\beta$  deposition in the grey matter was typical of that seen in Alzheimer's disease and A $\beta$  in the blood vessel walls was characteristic of cerebral amyloid angiopathy<sup>3</sup> and did not co-localize with prion protein deposition. None of these patients had pathogenic mutations, *APOE*  $\epsilon$ 4 or other high-risk alleles<sup>4</sup> associated with early-onset Alzheimer's disease. Examination of a series of 116 patients with other prion diseases from a prospective observational cohort study<sup>5</sup> showed minimal or no A $\beta$  pathology in cases of similar age range, or a decade older, without *APOE*  $\epsilon$ 4 risk alleles. We also analysed pituitary glands from individuals with A $\beta$  pathology and found marked A $\beta$  deposition in multiple cases. Experimental seeding of A $\beta$  pathology has been previously demonstrated in primates and transgenic mice by central nervous system or peripheral inoculation with Alzheimer's disease brain homogenate<sup>6–11</sup>. The marked deposition of parenchymal and vascular A $\beta$  in these relatively young patients with iCJD, in contrast with other prion disease patients and population controls, is consistent with iatrogenic transmission of A $\beta$  pathology in addition to CJD and suggests that healthy exposed individuals may also be at risk of iatrogenic Alzheimer's disease and cerebral amyloid angiopathy. These findings should also prompt investigation of whether other known iatrogenic routes of prion transmission may also be relevant to A $\beta$  and other proteopathic seeds associated with neurodegenerative and other human diseases.

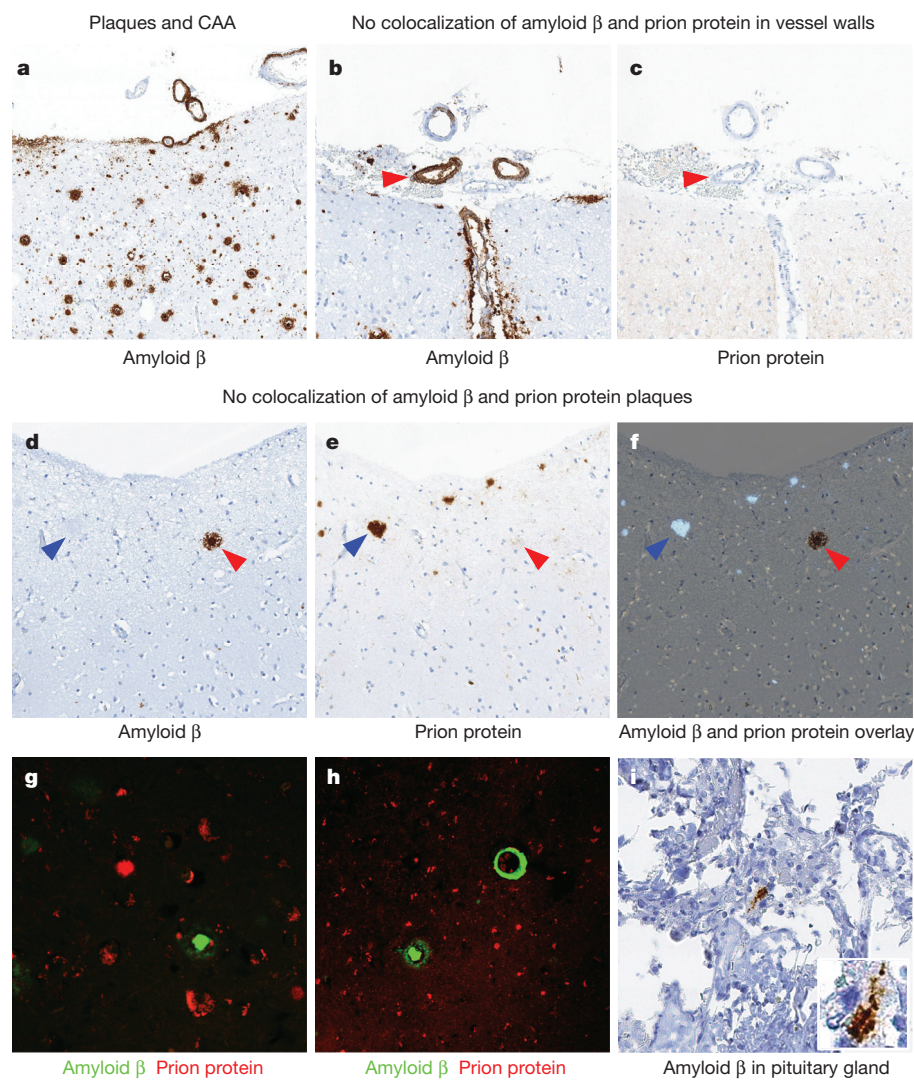
Human transmission of prion disease has occurred as a result of a range of medical and surgical procedures worldwide as well as by endocannibalism in Papua New Guinea, with incubation periods that can exceed five decades<sup>12,13</sup>. A well-recognized iatrogenic route of transmission was by treatment of persons of short stature with preparations of human growth hormone, extracted from large pools of cadaver-sourced pituitary glands, some of which were inadvertently prion-contaminated. Such treatments commenced in 1958 and ceased in 1985 following the reports of the occurrence of CJD amongst recipients. A review of all 1,848 patients who were treated with cadaveric-derived human growth hormone (c-hGH) in the United Kingdom from 1959 through 1985 found that 38 had developed CJD by the year 2000 with a peak incubation period of 20 years<sup>1</sup>. Multiple preparations using different extraction methods were used over this period and patients received batches from several preparations. One preparation

(Wilhelmi) was common to all patients who developed iCJD and size-exclusion chromatography, used in non-Wilhelmi preparation methods, may have reduced prion contamination<sup>1</sup>. As of 2012, a total of 450 cases of iatrogenic CJD have been recognized worldwide after treatment with c-hGH or gonadotropin (226 cases), transplantation of dura mater (228) or cornea (2), and neurosurgery (4) or electroencephalography recording using invasive medical devices (2)<sup>2</sup>. In France, 119/1,880 (6.3%) recipients developed iCJD, in the UK 65/1,800 (3.6%) and in the USA 29/7,700 (0.4%)<sup>2,14</sup>.

Since 2008, most UK patients with prion disease have been recruited into the National Prion Monitoring Cohort study<sup>5</sup>, including 22 of 24 recent patients with iatrogenic CJD (iCJD) related to treatment with c-hGH over this period, all of whom necessarily have very long incubation periods. Of this group of patients with iCJD, eight patients (referenced no.s 1–8, Supplementary Information) aged 36–51 years, with an incubation period from first treatment to onset of 27.9–38.9 years (mean 33 years) and from last treatment to onset of 18.8–30.8 years (mean 25.5 years), underwent autopsy with extensive brain tissue sampling at our hospital. In all eight brain samples we confirmed prion disease with abnormal prion protein labelling of the neuropil, perineuronal network and in most cases microplaques as described previously<sup>15–17</sup>. However, four (no.s 4, 5, 6, 8) of the eight patients with iCJD also showed substantial amyloid- $\beta$  (A $\beta$ ) deposition in the central nervous system parenchyma by histology (Fig. 1) and immunoblotting (Fig. 2). A further two brain samples (no.s 1, 3) had focal A $\beta$  pathology in one of the brain regions; one showed A $\beta$  entrapment in PrP plaques and only one was entirely negative for A $\beta$ . Furthermore, there was widespread cortical and leptomeningeal cerebral A $\beta$  angiopathy (CAA)<sup>3</sup> in three patients (no.s 4, 6, 8) and focal CAA in one patient (no. 5) (Fig. 1). Such pathology is extremely rare in this age range, 10/290 in the equivalent 36–50 year age strata without CJD<sup>18</sup>,  $P = 0.0002$ , Fisher's test. None of our patients with iCJD had pathogenic mutations in the prion protein gene (*PRNP*). We used a custom next generation sequencing panel<sup>4</sup> to exclude mutation in any of 16 other genes associated with early-onset Alzheimer's disease, CAA, or other neurodegenerative disorders, and none carried *APOE*  $\epsilon$ 4 or *TREM2* R47H alleles (Supplementary Table 2). Although such observations are unprecedented in our wide experience of human prion diseases, we nevertheless considered whether prion disease itself might predispose to, or accelerate, A $\beta$  pathology, for example by cross-seeding of protein aggregation or overload of clearance mechanisms for misfolded proteins. We therefore compared the A $\beta$  pathology in the iCJD cohort with that of a cohort of 116 patients with other prion diseases who had undergone autopsy: sporadic CJD (sCJD) ( $n = 85$ , age 42–83), variant CJD ( $n = 2$ , age 25 and 36) and inherited prion diseases (IPD) ( $n = 29$ , age 29–86). None of the patients in the control cohorts had comparable A $\beta$  pathology (Consortium to Establish a Registry for Alzheimer's disease (CERAD) score,  $P = 0.001$ , CAA,

<sup>1</sup>Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK. <sup>2</sup>Medical Research Council Prion Unit, Queen Square, London WC1N 3BG, UK.

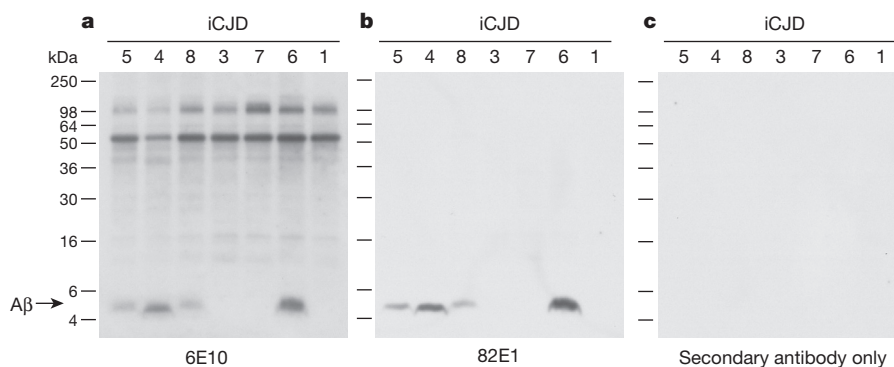
<sup>3</sup>Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK. <sup>4</sup>National Prion Clinic, The National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK. <sup>5</sup>MRC Clinical Trials Unit at University College London, 125 Kingsway, London WC2B 6NH, UK.



**Figure 1 | A $\beta$  accumulation in central nervous system parenchyma and blood vessels (CAA) in iCJD.** **a**, Frontal cortex with widespread diffuse A $\beta$  deposition, formation of plaques, and widespread parenchymal and leptomeningeal CAA (patient no. 4). **b**, **c**, Non-colocalized deposition of A $\beta$  and prion protein. Vessels with CAA do not entrap or co-seed prion protein. **d**, **e**, Adjacent histological sections stained for A $\beta$  or prion protein show clearly separated plaques of both proteins (no. 5). **f**, An overlay with colour inversion of prion protein plaques highlights the separation. **g**, **h**, Dual labelling, confocal laser microscopy shows no co-localization of parenchymal A $\beta$  plaques (no.s 5, 6) or CAA (no. 6). **i**, A $\beta$  is detected in pituitary glands in patients with a high A $\beta$  load in the brain. Scale bar corresponds to 200  $\mu$ m in **a**, 100  $\mu$ m in **b–h**, and 50  $\mu$ m in **i**.

$P = 0.005$ , topographical A $\beta$  score  $P = 0.02$ , and cumulative A $\beta$  score  $P = 0.02$  (rank sum test) and digital A $\beta$  quantification  $P = 0.04$  ( $t$ -test); all restricted to the strata aged 36–51 years ( $n = 19$ )) (Fig. 3, and Extended Data Figs 1 and 2 show similar results in adjusted analyses in the full cohort). Indeed none of 35 prion cases aged 52–60 had significant A $\beta$  pathology, with the exception of two cases at ages 57

and 58 positive for *APOE*  $\epsilon 4$  alleles. Instead, the sCJD cohort shows A $\beta$  pathology in parenchyma and blood vessels to a similar extent/severity as seen in iCJD, only in a much older age group (Extended Data Figs 1 and 2), in keeping with the chance coincidence of late-onset A $\beta$  pathology and sCJD as previously documented in a large study of 110 sCJD patients and 110 age-matched controls aged 27–84 (ref. 19) and a study



**Figure 2 | Immunoblots of A $\beta$  in iCJD patient brains.** **a–c**, 10% (w/v) brain homogenates from patients with iCJD were analysed by enhanced chemiluminescence using anti-human A $\beta$  monoclonal antibodies 6E10 that recognizes full-length APP and fragments that contain the epitope including A $\beta$  (**a**) or 82E1 that specifically recognizes A $\beta$  (**b**) or secondary antibody only

(**c**). The identity of the patient brain sample is designated above each lane and the position of molecular mass markers is shown to the left. The equivalent of 5  $\mu$ l 10% (w/v) brain homogenate was loaded per lane. The migration position of A $\beta$  is indicated by the arrow. For gel source data, see Supplementary Fig. 1.



of 2,661 individuals aged 26–95 (ref. 18). Further, we investigated whether prion and A $\beta$  pathology co-localize in the iCJD cases. In our series there was a distinct absence of overlap of A $\beta$  plaques and PrP (Fig. 1d, e) or A $\beta$  CAA and vascular PrP (Fig. 1b, c), consistent with these pathologies developing independently.

We then went on to examine pituitary glands for the presence of A $\beta$  deposits. Pathological species of tau, A $\beta$  and  $\alpha$ -synuclein have been reported in the pituitary gland of patients with neurodegenerative disease and controls<sup>20</sup>. We examined 55 pituitary glands, 6 from patients without, and 49 from patients with cerebral A $\beta$  pathology, and found in the latter group seven samples containing A $\beta$ , confirming frequent A $\beta$  in pituitaries of patients with Alzheimer's disease-like pathology<sup>20</sup> (Fig. 1i and Extended Data Fig. 3), consistent with the hypothesis that A $\beta$  seeds have been iatrogenically transmitted to these patients with iCJD.

There has been longstanding interest as to whether other neurodegenerative diseases associated with the accumulation of aggregates of misfolded host proteins or amyloids might be transmissible in a 'prion-like' fashion<sup>21,22</sup>. Experimental seeding of A $\beta$  pathology has previously been demonstrated in primates and transgenic mice by central nervous system inoculation with Alzheimer's disease brain homogenate<sup>6–10</sup>. Of particular interest with respect to our findings is that peripheral (intra-peritoneal) inoculation with Alzheimer's disease brain extract into APP23 (ref. 11) transgenic mice has been demonstrated. While ageing APP23 mice show mostly parenchymal deposits, the intraperitoneally-seeded mice showed predominantly CAA, a feature seen in patients with iCJD who had significant A $\beta$  pathology. This experimental study and our findings suggest that there are mechanisms to allow the trans-

port of A $\beta$  seeds as well as prions (and possibly other proteopathic seeds such as tau<sup>23</sup>) from the periphery to the brain<sup>24,25</sup>. While less than 4% of UK c-hGH treated individuals have developed iCJD, one out of eight patients with iCJD had focal, and three had widespread, moderate or severe CAA. Four patients had widespread parenchymal A $\beta$  pathology and two further patients had focal cortical A $\beta$  deposits. This might suggest that healthy individuals exposed to c-hGH are at high risk of developing early-onset A $\beta$  pathology as this cohort ages.

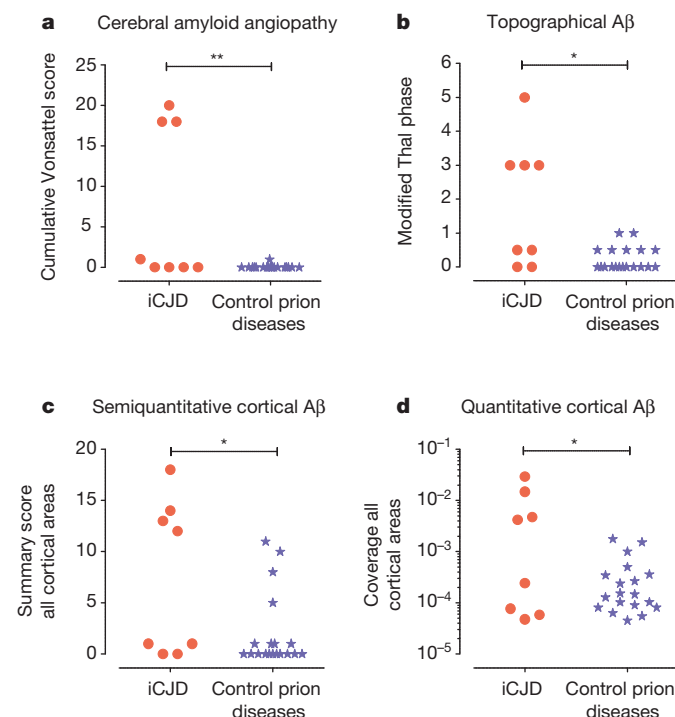
Although none of the iCJD cases with A $\beta$  pathology had hyperphosphorylated tau neurofibrillary tangle pathology characteristic of Alzheimer's disease, it is possible that the full neuropathology of Alzheimer's disease would have developed had these individuals not succumbed to prion disease at these relatively young ages. An earlier study concluded that c-hGH recipients did not seem to be at increased risk of Alzheimer's disease, but this was based on death certificates only without autopsy data<sup>20</sup>. However, the severe CAA seen in the patients with iCJD in our study is unquestionably concerning and individuals with such pathology would be at increasing risk of cerebral haemorrhages had they lived longer. At-risk individuals, including patients who had received dura mater grafts<sup>26</sup> could be screened by magnetic resonance imaging (MRI) for CAA-related pathologies (such as microbleeds) and by positron emission tomography (PET) for A $\beta$  deposition<sup>27</sup>.

It is possible, however, that prions and A $\beta$  seeds co-purify in the extraction methods used to prepare c-hGH, which might mean that there would be a relatively higher occurrence of A $\beta$  pathology in those with iatrogenic prion infection. Analysis of any residual archival batches of c-hGH for both prions and A $\beta$  seeds might be informative in this regard<sup>2</sup>. While our data argue against cross seeding, we cannot formally exclude the possibility that prions somehow seed A $\beta$  deposition but do not co-localize with A $\beta$  deposits. While there is no suggestion that Alzheimer's disease is a contagious disease and no supportive evidence from epidemiological studies that Alzheimer's disease is transmissible, notably by blood transfusion<sup>28,29</sup>, our findings should prompt consideration of whether other known iatrogenic routes of prion transmission, including surgical instruments and blood products, may also be relevant to A $\beta$  and other proteopathic seeds seen in neurodegenerative diseases. A $\beta$  seeds are known, like prions, to adhere to metal surfaces and to resist formaldehyde inactivation and conventional hospital sterilisation<sup>30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 26 April; accepted 14 August 2015.**

1. Swerdlow, A. J., Higgins, C. D., Adlard, P., Jones, M. E. & Preece, M. A. Creutzfeldt-Jakob disease in United Kingdom patients treated with human pituitary growth hormone. *Neurology* **61**, 783–791 (2003).
2. Brown, P. et al. Iatrogenic Creutzfeldt-Jakob disease, final assessment. *Emerg. Infect. Dis.* **18**, 901–907 (2012).
3. Vonsattel, J. P. et al. Cerebral amyloid angiopathy without and with cerebral hemorrhages: a comparative histological study. *Ann. Neurol.* **30**, 637–649 (1991).
4. Beck, J. et al. Validation of next-generation sequencing technologies in genetic diagnosis of dementia. *Neurobiol. Aging* **35**, 261–265 (2014).
5. Rudge, P. et al. Iatrogenic CJD due to pituitary-derived growth hormone with genetically determined incubation times of up to 40 years. *Brain*. <http://dx.doi.org/10.1093/brain/awv235> (2015).
6. Baker, H. F., Ridley, R. M., Duchon, L. W., Crow, T. J. & Bruton, C. J. Induction of beta (A4)-amyloid in primates by injection of Alzheimer's disease brain homogenate. Comparison with transmission of spongiform encephalopathy. *Mol. Neurobiol.* **8**, 25–39 (1994).
7. Eisele, Y. S. et al. Induction of cerebral  $\beta$ -amyloidosis: intracerebral versus systemic A $\beta$  inoculation. *Proc. Natl Acad. Sci. USA* **106**, 12926–12931 (2009).
8. Hamaguchi, T. et al. The presence of A $\beta$  seeds, and not age per se, is critical to the initiation of A $\beta$  deposition in the brain. *Acta Neuropathol.* **123**, 31–37 (2012).
9. Heilbronner, G. et al. Seeded strain-like transmission of  $\beta$ -amyloid morphotypes in APP transgenic mice. *EMBO Rep.* **14**, 1017–1022 (2013).
10. Meyer-Luehmann, M. et al. Exogenous induction of cerebral  $\beta$ -amyloidogenesis is governed by agent and host. *Science* **313**, 1781–1784 (2006).
11. Eisele, Y. S. et al. Peripherally applied A $\beta$ -containing inoculates induce cerebral  $\beta$ -amyloidosis. *Science* **330**, 980–982 (2010).



**Figure 3 | Early A $\beta$  accumulation in the parenchyma and blood vessels in a subset of eight patients with iCJD aged 36–51 years, but not in controls (stratum aged 36–51 years) of 19 prion diseases of other aetiologies, suggests human transmission. a**, Widespread, moderate-to-severe early-onset CAA in three, and focal, mild CAA in one iCJD patient but only one focal, mild CAA in 19 controls. **b**, Significant differences of parenchymal A $\beta$  accumulation (all central nervous system regions, see supplementary material). **c**, **d**, Cortical A $\beta$  load was assessed semiquantitatively and quantitatively and again was significantly different between the iCJD and age-matched control cohort. For methods of quantification and calculations of significance levels see Supplementary Information.

12. Collinge, J. Prion diseases of humans and animals: their causes and molecular basis. *Annu Rev. Neurosci.* **24**, 519–550 (2001).
13. Collinge, J. *et al.* Kuru in the 21st century—an acquired human prion disease with very long incubation periods. *Lancet* **367**, 2068–2074 (2006).
14. Abrams, J. Y. *et al.* Lower risk of Creutzfeldt-Jakob disease in pituitary growth hormone recipients initiating treatment after 1977. *J. Clin. Endocrinol. Metab.* **96**, E1666–E1669 (2011).
15. Brandner, S. *et al.* Central and peripheral pathology of kuru: pathological analysis of a recent case and comparison with other forms of human prion disease. *Phil. Trans. R. Soc. Lond. B* **363**, 3755–3763 (2008).
16. Ellison, D. *et al.* *Neuropathology: A reference Text of CNS Pathology*, 3rd edn (Elsevier, 2012).
17. Parchi, P. *et al.* Consensus classification of human prion disease histotypes allows reliable identification of molecular subtypes: an inter-rater study among surveillance centres in Europe and USA. *Acta Neuropathol.* **124**, 517–529 (2012).
18. Braak, H. & Braak, E. Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiol. Aging* **18**, 351–357 (1997).
19. Hainfellner, J. A. *et al.* Coexistence of Alzheimer-type neuropathology in Creutzfeldt-Jakob disease. *Acta Neuropathol.* **96**, 116–122 (1998).
20. Irwin, D. J. *et al.* Evaluation of potential infectivity of Alzheimer and Parkinson disease proteins in recipients of cadaver-derived human growth hormone. *JAMA Neurol.* **70**, 462–468 (2013).
21. Collinge, J. & Clarke, A. R. A general model of prion strains and their pathogenicity. *Science* **318**, 930–936 (2007).
22. Gajdusek, D. C. Transmissible and non-transmissible amyloidoses: autocatalytic post-translational conversion of host precursor proteins to beta-pleated sheet configurations. *J. Neuroimmunol.* **20**, 95–110 (1988).
23. Clavaguera, F. *et al.* Peripheral administration of tau aggregates triggers intracerebral tauopathy in transgenic mice. *Acta Neuropathol.* **127**, 299–301 (2014).
24. Beekes, M., Thomzig, A., Schulz-Schaeffer, W. J. & Burger, R. Is there a risk of prion-like disease transmission by Alzheimer- or Parkinson-associated protein particles? *Acta Neuropathol.* **128**, 463–476 (2014).
25. Bolmont, T. *et al.* Induction of tau pathology by intracerebral infusion of amyloid- $\beta$ -containing brain extract and by amyloid-beta deposition in APP  $\times$  Tau transgenic mice. *Am. J. Pathol.* **171**, 2012–2020 (2007).
26. Preusser, M. *et al.* Alzheimer-type neuropathology in a 28 year old patient with iatrogenic Creutzfeldt-Jakob disease after dural grafting. *J. Neurol. Neurosurg. Psychiatry* **77**, 413–416 (2006).
27. Rowe, C. C. *et al.* Imaging of amyloid beta in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: proof of mechanism. *Lancet Neurol.* **7**, 129–135 (2008).
28. Daviglus, M. L. *et al.* Risk factors and preventive interventions for Alzheimer disease: state of the science. *Arch. Neurol.* **68**, 1185–1190 (2011).
29. O'Meara, E. S. *et al.* Alzheimer's disease and history of blood transfusion by apolipoprotein-E genotype. *Neuroepidemiology* **16**, 86–93 (1997).
30. Fritsch, S. K. *et al.* A $\beta$  seeds resist inactivation by formaldehyde. *Acta Neuropathol.* **128**, 477–484 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was funded by the UK Medical Research Council and the National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre and Dementia Biomedical Research Unit. We are grateful to all patients and their relatives and carers for their participation in the National Prion Monitoring Cohort and for consent to autopsy and use of tissues for this research. We thank all physicians who contributed information and the National CJD Research and Surveillance Unit for coordination of patient referral and tissue sharing under the National referral scheme. We also thank J. Broni, L. Brock and T. Wilkins for histological assistance and B. Peters, R. Sinclair, C.-W. Lok and M. Karseras for support with post mortem examinations at UCLH. We thank G. Adamson, T. Campbell, J. Uphill and R. Drueyeh for assistance with genetic analyses.

**Author Contributions** Z.J. and S.B. performed autopsies and neuropathological diagnosis and analysis of cases. J.K., S.M., P.R. and J.C. diagnosed and recruited patients to the study and analysed clinical and genetic data. A.S.W. performed statistical analysis and modelling. J.D.F.W. and A.J.N. performed western blots, M.E., F.L., A.R.-L. and J.L. performed immunohistochemistry and analysed data, J.C. and S.B. oversaw the study and drafted the manuscript with contributions from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.B. ([s.brandner@ucl.ac.uk](mailto:s.brandner@ucl.ac.uk)) or J.C. ([j.collinge@prion.ucl.ac.uk](mailto:j.collinge@prion.ucl.ac.uk)).

## METHODS

No statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Patient recruitment and genotyping.** A national referral system for prion diseases was established by the Chief Medical Officer in the UK in 2004. UK neurologists were asked to refer all patients with suspected prion disease jointly to the National CJD Research and Surveillance Unit in Edinburgh and the NHS National Prion Clinic (NPC) in London. All patients with possible CJD who had received cadaver-derived growth hormone were referred to the NHS National Prion Clinic (London, UK) and since 2008 were recruited into the National Prion Monitoring Cohort study.

**Next-generation sequencing to exclude mutations known to be causal of A $\beta$  pathology.** Deep next-generation sequencing using a custom panel was performed as described previously<sup>31</sup>. Analysis was done using NextGENe and Geneticist Assistant software (Softgenetics, USA). Variants were assessed for pathogenicity by reference to the published literature, control population allele frequencies (our primary database for allele frequency was the Broad Institute's ExAC browser (<http://exac.broadinstitute.org/>)) and *in silico* predictive tools. The analysis methodology has been validated for the detection of APP duplication<sup>31</sup>, which was important to exclude. No causal mutations for dementia or A $\beta$  pathology were detected, see Supplementary Table 2. As expected, several rare variants were detected which may modify the risk of various neurodegenerative diseases, see Supplementary Table 2.

**Autopsies and tissue preparation.** Autopsies were carried out in a post mortem room designated for high risk autopsies. Informed consent to use the tissue for research was obtained in all cases. Ethical approval for these studies was obtained from the Local Research Ethics Committee of the UCL Institute of Neurology/National Hospital for Neurology and Neurosurgery. The anterior frontal, temporal, parietal and occipital cortex and the cerebellum (at the level of dentate nucleus) were dissected during the post mortem procedure and frozen. Samples of the following areas were taken and analysed: frontal, temporal, parietal, occipital, posterior frontal cortex including motor strip, basal ganglia, thalamus, hippocampus, brain stem including midbrain, and cerebellar hemisphere and vermis. Pituitary glands were taken in all cases.

Tissue samples were immersed in 10% buffered formalin and prion infectivity was inactivated by immersion into 98% formic acid for one hour. Tissue samples were processed to paraffin wax and tissue sections were routinely stained with haematoxylin and eosin.

**Antibodies and immunohistochemistry.** The following antibodies were used: Anti-PrP ICSM35 (D-Gen Ltd, London, UK<sup>32,33</sup> 1:1,000), Anti-phospho-Tau (AT-8, Innogenetics, 1:100) and anti- $\beta$ A4 (DAKO 6F3D, 1:50). ICSM35 was stained on a Ventana Benchmark or Discovery automated immunohistochemical staining machine (ROCHE Burgess Hill, UK);  $\beta$ A4 and Tau were stained on a LEICA BondMax (LEICA Microsystems) or a Ventana automated staining instrument following the manufacturer's guidelines, using biotinylated secondary antibodies and a horseradish-peroxidase-conjugated streptavidin complex and diaminobenzidine as a chromogen.

**Immunoblot detection of A $\beta$  in iCJD brain.** Biochemical studies were carried out in a microbiological containment level 3 facility with strict adherence to safety protocols. Frozen brain tissue was available from seven of eight patients with growth hormone iCJD (cases 1 and 3–8). 10% (w/v) brain homogenates (grey matter; frontal cortex) were prepared in Dulbecco's PBS lacking Ca<sup>2+</sup> or Mg<sup>2+</sup> ions using tissue grinders as described previously<sup>34</sup>. 20- $\mu$ l aliquots were treated with 1  $\mu$ l benzamide nuclease (purity >99%; 25 U ml<sup>-1</sup>; Novagen) for 15 min at 20 °C. Samples were then mixed with an equal volume of 2 $\times$  SDS sample buffer (125 mM Tris-HCl, 20% (v/v) glycerol pH 6.8 containing 4% (w/v) SDS, 4% (v/v) 2-mercaptoethanol and 0.02% (w/v) bromophenol blue) and immediately transferred to a 100 °C heating block for 10 min. Electrophoresis was performed on 16% Tris-glycine gels (Invitrogen), run for 70 min at 200 V, before electroblotting to Immobilon P membrane (Millipore) for 16 h at 15 V as described previously<sup>34</sup>. Membranes were blocked in phosphate buffered saline (PBS) containing 0.05% (v/v) Tween 20 (PBST) and 5% (w/v) non-fat dried skimmed milk powder. Blots were then probed with anti-human A $\beta$  monoclonal antibodies 6E10 (Covance) and 82E1 (IBL international, Hamburg, Germany) at final concentrations of 0.2  $\mu$ g ml<sup>-1</sup> in PBST for at least 1 h. After washing for 1 h with PBST the membranes were probed with a 1:10,000 dilution of alkaline-phosphatase-conjugated goat anti-mouse IgG secondary antibody (Sigma-Aldrich no. A2179) in PBST. After washing (90 min with PBST and 5 min with 20 mM Tris pH 9.8 containing 1 mM MgCl<sub>2</sub>) blots were incubated for 5 min in chemiluminescent substrate (CDP-Star; Tropix Inc.) and visualized on Biomax MR film (Carestream Health Inc.). Anti-human A $\beta$  monoclonal antibody 82E1 recognizes an epitope specific to the amino terminus of A $\beta$  while 6E10 recognizes an epitope spanning residues 3–8 of A $\beta$  and cross-reacts with full-length APP or APP fragments that contain the epitope.

**Examination of prion pathology.** In all iCJD cases there was variably prominent microvacuolar change in the neocortices, deep grey nuclei and cerebellar cortex. Immunostaining for the abnormal prion protein revealed synaptic labelling in all grey matter areas examined. In all but one case, there were also microplaques in all grey matter structures. Variability in the intensity of the immunoreactivity for the abnormal prion protein was evident but detailed comparison between the cases and separately within each case was not feasible as prolonged formalin fixation in some cases significantly attenuated the immunoreactivity. It was apparent that more prominent microvacuolar change and synaptic labelling for abnormal prion protein was more intense in the pre-central gyrus and parietal lobe when compared to the anterior frontal and occipital cortices. Deep cortical layers showed more severe changes. In all cases the microvacuolar degeneration and prion protein deposits in the deep grey nuclei and hippocampal formation was prominent. It was most severe in the caudate nucleus and putamen, and appeared less severe in thalamus and it was least prominent in the globus pallidus. In the cerebellar vermis there was marked granule cell atrophy and often widespread loss of Purkinje cells accompanied by severe Bergmann gliosis, while cerebellar hemispherical cortex showed only patchy loss of Purkinje cells and no significant granule cell loss. Microvacuolar degeneration in the molecular layer was more prominent in the vermis than in the cerebellar hemisphere. No apparent difference in prion protein deposition was seen in vermis and hemisphere. In the dentate nucleus variably intense synaptic prion protein immunoreactivity was present, while the cyto-architecture of the nucleus was well preserved.

**Examination, classification and quantification of A $\beta$  pathology.** All brains were examined according to the ABC classification<sup>35</sup>, which assesses the topographic progression of A $\beta$  pathology in the brain (Thal phases<sup>36</sup>), topographic progression of Tau neurofibrillary tangle pathology (Braak and Braak<sup>37</sup>) and the density of mature (senile), neuritic plaques in the neocortex (Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria<sup>38,39</sup>). To allow a more detailed assessment of neocortical A $\beta$  the original Thal phases were modified as follows. Phase 0, no cortical A $\beta$ ; phase 0.5, 1–2 neocortical regions affected; phase 1, 3–4 neocortical regions involved; phases 2–5 were scored as published<sup>36</sup>. In addition we have carried out a semiquantitative assessment of neocortical A $\beta$  load on a standardised region within frontal, temporal, parietal and occipital lobes, and scored as follows. 0, entirely negative; 1, a single small deposit; 2, multiple small deposits, disseminated; 3, multiple small deposits, plus an area with a larger patch; 4, diffuse moderate numbers of deposits; 5, diffuse, frequent numbers of deposits. For each case a cumulative score (0–20) of total semiquantitatively assessed A $\beta$  load in the neocortex was calculated. Cerebral amyloid angiopathy (CAA) was graded (0–3) according to the Vonsattel criteria<sup>3</sup>. CAA was assessed in leptomeninges and parenchyma of all hemispheric lobes and cerebellum with summary score (0–30) calculated for each case.

**Image acquisition and processing.** Histological slides were digitised on a LEICA SCN400F scanner (LEICA Milton Keynes, UK) at  $\times 40$  magnification and 65% image compression setting during export. Slides were archived and managed on LEICA Slidepath (LEICA Milton Keynes, UK). For the preparation of light microscopy images, 1,024  $\times$  1,024 pixel sized image captures were taken, after matching paired images (A $\beta$  and prion staining) in Slidepath, and overlays in Fig. 1f were prepared using the colour conversion function in conjunction with the image overlay in Slidepath. Laser scanning microscopy of double immunofluorescent tissue preparations was on a ZEISS LSM710 confocal microscope (ZEISS Cambridge, UK). Publication figures were assembled in Adobe Photoshop. Data plots were generated using Prism 5 (GraphPad Software, Inc., La Jolla, USA).

**Digital image analysis for cortical A $\beta$  quantification.** From all cases A $\beta$  immunostained slides from frontal, temporal, parietal and occipital lobes were digitised as described above. Digital image analysis on 496 whole slides was performed using Definiens Developer 2.3 (Definiens, Munich, Germany). Initial tissue identification was performed at a resolution corresponding to 5 $\times$  image magnification and stain detection was performed at  $\times 10$  resolution. Tissue detection and initial segmentation was done to identify all tissue within the image, separating the sample from background and non-tissue regions for further analysis. This separation was based on identification of the highly homologous relatively bright/white region of background present at the perimeter of each image. A composite raster image produced by selecting the lowest pixel value from the three comprising colour layers (RGB colour model) provided a greyscale representation of brightness. The mean brightness of this background region was used to exclude all background regions from further analysis.

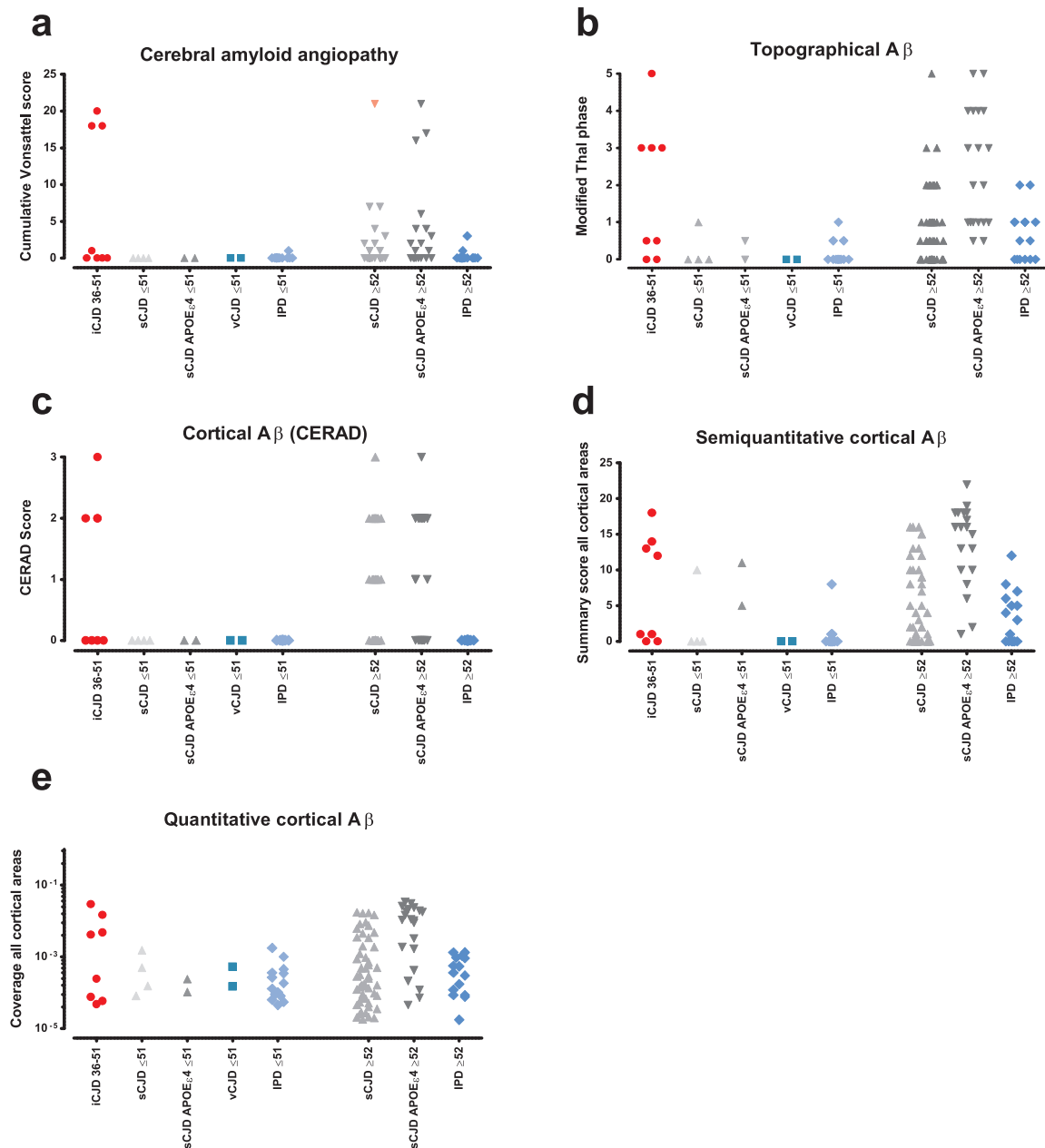
Stain detection (brown) is based on the transformation of the RGB colour model to a HSD representation<sup>40</sup>. This provides a raster image of the intensity of each colour of interest (brown and blue). A series of dynamic thresholds ( $T_x$ ) are then used to identify areas of interest ( $A_x$ ). Initially, following exclusion of intensely stained areas with values greater than 1 arbitrary unit (au) (values range from 0au to 3au in HSD images), the 5th centile ( $C_5$ ) of brown stain intensity was calculated



as a baseline. This represents the  $T_{\text{brown stain}}$  separating the top 5% of  $A_{\text{tissue}}$ . The standard deviation ( $C5\delta$ ) within the lower 95% of  $A_{\text{tissue}}$  was used to update the  $T_{\text{brown stain}}$  as  $\overline{C5} + (6 \times C5\delta)$  with all pixels above this threshold classed as 'stain' ( $A_{\text{stain}}$ ) and those below as 'unstained' ( $A_{\text{unstained}}$ ).  $A_{\text{stain}}$  was excluded if the intensity of blue staining was not significantly lower than the level of brown stain (difference less than 0.1au) to remove generically dark areas. The remaining  $A_{\text{stain}}$  were further categorised using thresholds based on the mean ( $\bar{B}$ ) and standard deviation ( $B\delta$ ) of brown staining within the  $A_{\text{unstained}}$ :  $T_{\text{brown}} = \bar{B} + (3 \times B\delta)$  (lower threshold);  $T_{\text{dark brown}} = \bar{B} + (6 \times B\delta)$  (upper threshold), to give  $A_{\text{unstained}} \leq T_{\text{brown}} > A_{\text{light brown}} \leq T_{\text{dark brown}} > A_{\text{A}\beta \text{ deposit}}$ . Artefacts were then identified as  $A_{\text{stain}}$  with area greater than  $1 \text{ mm}^2$ , or an area greater than  $0.1 \text{ mm}^2$  with a standard deviation of brown staining below 0.2au. These  $A_{\text{artefacts}}$  were then expanded to include surrounding pixels with brown staining greater than  $\overline{C5}$ . This excludes large areas of homogenous staining and areas of more diffuse, non-specific chromogen deposit.

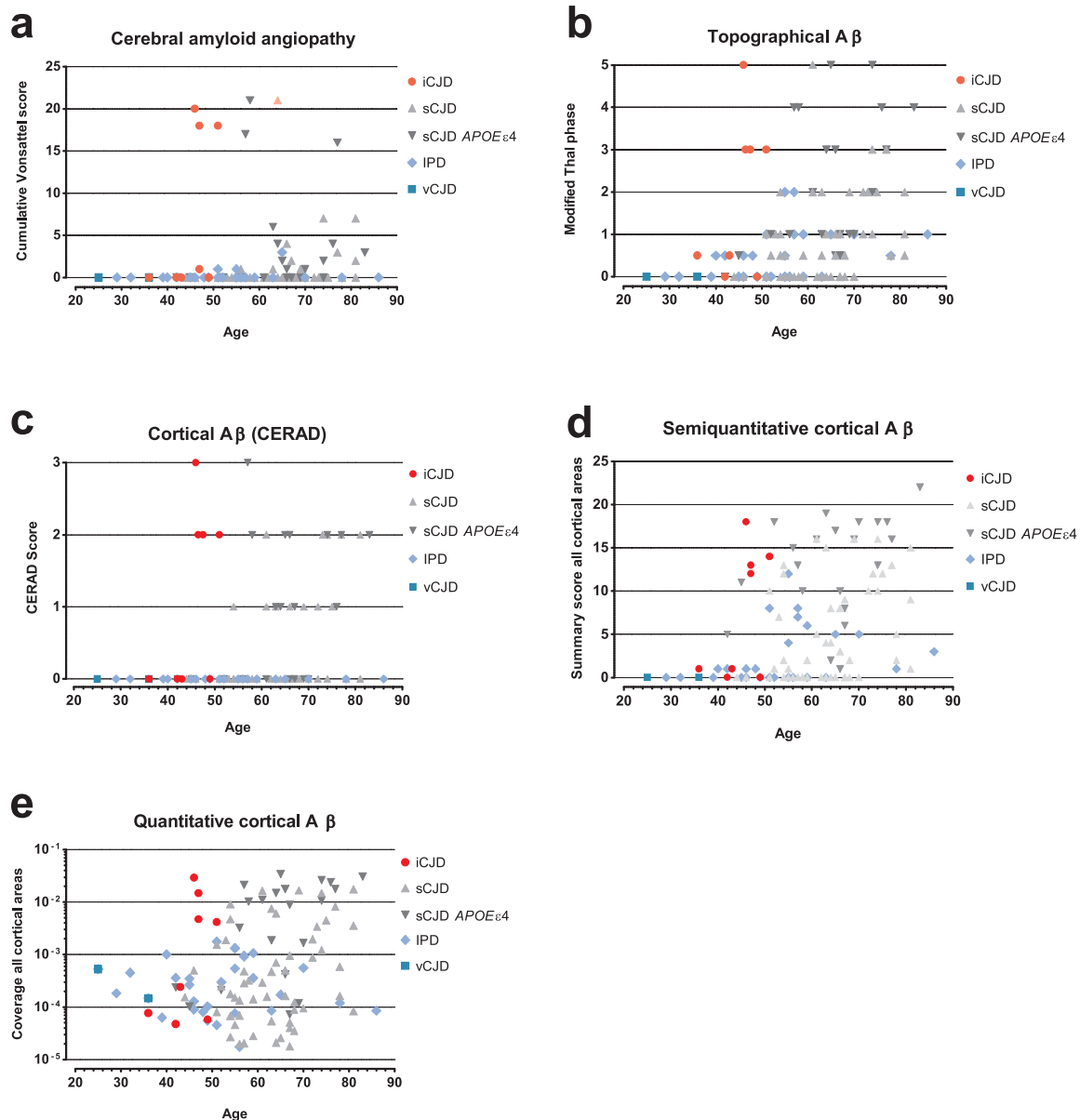
The white matter region within the tissue was then manually selected by an expert neuropathologist (Z.J., S.B.). This white matter was excluded from calculation of proportional coverage of  $A_{\text{A}\beta \text{ deposit}}$  within  $A_{\text{tissue}}$ .

31. Beck, J. *et al.* Validation of next-generation sequencing technologies in genetic diagnosis of dementia. *Neurobiol. Aging* **35**, 261–265 (2014).
32. Isaacs, A. M. *et al.* Lack of TAR-DNA binding protein-43 (TDP-43) pathology in human prion diseases. *Neuropathol. Appl. Neurobiol.* **34**, 446–456 (2008).
33. Khalili-Shirazi, A. *et al.* PrP glycoforms are associated in a strain-specific ratio in native PrPSc. *J. Gen. Virol.* **86**, 2635–2644 (2005).
34. Wadsworth, J. D. *et al.* Molecular diagnosis of human prion disease. *Methods Mol. Biol.* **459**, 197–227 (2008).
35. Hyman, B. T. *et al.* National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* **8**, 1–13 (2012).
36. Thal, D. R., Rüb, U., Orantes, M. & Braak, H. Phases of A $\beta$ -deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791–1800 (2002).
37. Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112**, 389–404 (2006).
38. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).
39. Heyman, A., Fillenbaum, G. G. & Mirra, S. S. Consortium to Establish a Registry for Alzheimer's Disease (CERAD): clinical, neuropsychological, and neuropathological components. *Aging (Milano)* **2**, 415–424 (1990).
40. van Der Laak, J. A., Pahlplatz, M. M., Hanselaar, A. G. & de Wilde, P. C. Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy. *Cytometry* **39**, 275–284 (2000).



**Extended Data Figure 1 | Vertical scatter plot of A $\beta$  pathology in iCJD brain samples, age-matched and older controls.** **a**, CAA summary score as described in Methods. iCJD (age range 36–51) with three highly scoring brain samples with CAA. In the age-matched control groups, no comparable pathology was found. Significant CAA was only seen in a cohort including older individuals carrying the APOE  $\epsilon_4$  allele. The outlier in the sCJD group (pink triangle in the sCJD  $> 52$  year group) had a surgical intervention 40 years before death, and in addition to CJD also had severe CAA. **b**, Topographical A $\beta$  deposition, assessed according to a scheme modified from the Thal classification as described in the Methods section. In the group of individuals of 51 years

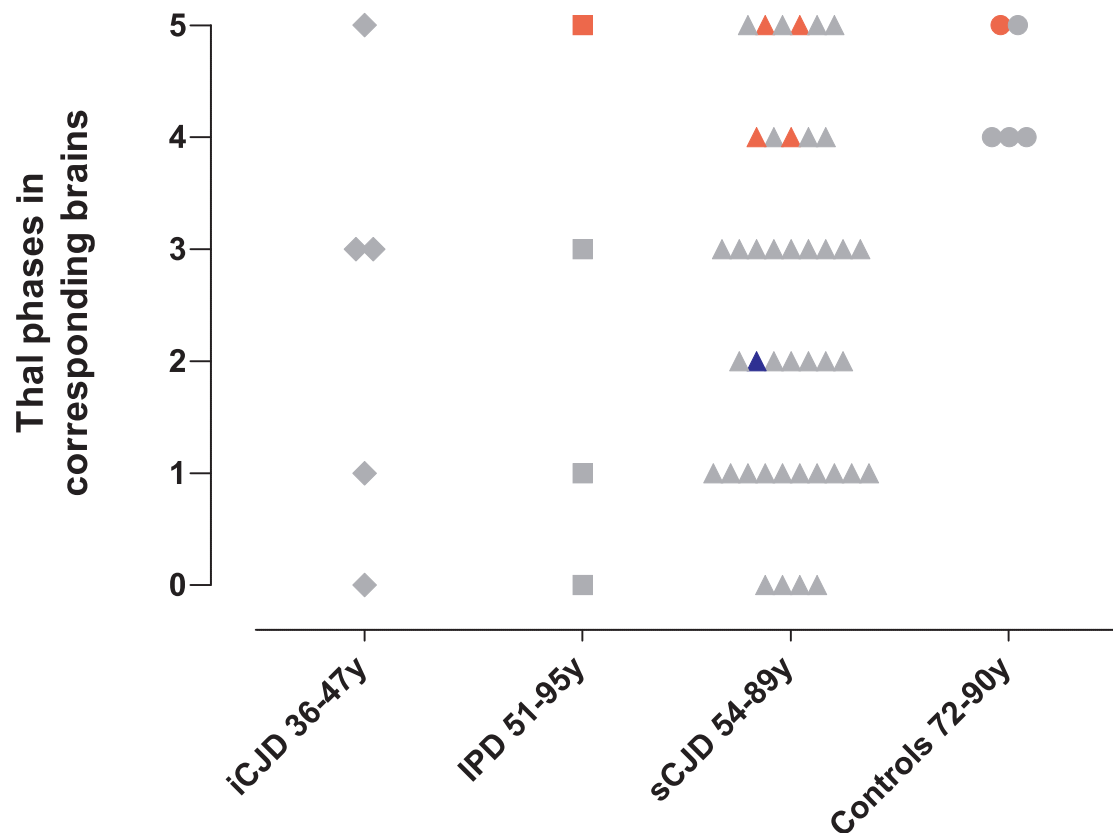
and below, significant A $\beta$  deposition is seen in the iCJD group, but not in age-matched prion diseases of different aetiology. **c**, Cortical A $\beta$  deposition, assessed according to CERAD. In the group of individuals of 51 years and below, mature (neuritic) plaques are seen in the iCJD group, but not in age-matched prion diseases of different aetiology. Only in the cohort comprising much older individuals is there an increase of cortical mature plaques. **d**, **e**, Semiquantitative assessment and quantification of neocortical A $\beta$  using Definiens Developer image analysis shows a separation that is similar to that shown in **a** and **b**. APOE  $\epsilon_4$  genotype was unavailable in nine of the 85 patients with sCJD and these were not included in the graphs. Note the logarithmic scale in **e**.



**Extended Data Figure 2 | Scattergram of A $\beta$  pathology in iCJD brain samples compared with other prion diseases.** Plot of severity scores of CAA or parenchymal A $\beta$  against the age of individuals in the cohorts, demonstrating early-onset of CAA and grey matter A $\beta$  pathology in the iCJD cohort. **a**, Early-onset of CAA. The outlier in the sCJD group (pink triangle) had a surgical intervention 40 years before death, and in addition to CJD also had

severe CAA. **b, c**, Early detection of grey matter A $\beta$  by a topographical assessment (Thal phase) and using CERAD Criteria. **d, e**, Semiquantitative assessment and quantification of neocortical A $\beta$ . *APOE*  $\epsilon$ 4 genotype was unavailable in nine of the 85 patients with sCJD and these were not included in the graph. Note the logarithmic scale in **e**.





**Extended Data Figure 3 | Topographical A $\beta$  of the brain samples corresponding to the pituitaries analysed for A $\beta$ .** Plot of the Thal phases (topographical A $\beta$  deposition) of the brain samples corresponding to the 55 pituitary glands examined for the presence of A $\beta$ . A $\beta$  was assessed in 55 pituitary glands from patients with iCJD ( $n = 5$ , age range 36–47), IPD ( $n = 4$ , age range 51–95), sCJD ( $n = 41$ , age range 54–89) and non-CJD controls ( $n = 5$ , age range 72–90) (groups shown on  $x$  axis). In six patients from iCJD, IPD and sCJD groups no A $\beta$  deposits were found in the brain or pituitary gland.

In 49 patients from all groups (iCJD, IPD, sCJD and non-CJD) there were variably frequent A $\beta$  deposits in the brain parenchyma, corresponding to Thal phases 1–5 (distribution shown on  $y$  axis). Of these 49 cases, six cases (IPD  $n = 1$ , sCJD  $n = 4$ , and non-CJD  $n = 1$ ) showed A $\beta$  deposits also in the pituitary glands (positive cases highlighted in red) and in one patient from sCJD group A $\beta$  deposits were seen in the brain tissue attached to the pituitary gland (highlighted in blue).

# Single-cell messenger RNA sequencing reveals rare intestinal cell types

Dominic Grün<sup>1,2\*</sup>, Anna Lyubimova<sup>1,2\*</sup>, Lennart Kester<sup>1,2</sup>, Kay Wiebrands<sup>1,2</sup>, Onur Basak<sup>1,2</sup>, Nobuo Sasaki<sup>1,2</sup>, Hans Clevers<sup>1,2</sup> & Alexander van Oudenaarden<sup>1,2</sup>

Understanding the development and function of an organ requires the characterization of all of its cell types. Traditional methods for visualizing and isolating subpopulations of cells are based on messenger RNA or protein expression of only a few known marker genes. The unequivocal identification of a specific marker gene, however, poses a major challenge, particularly if this cell type is rare. Identifying rare cell types, such as stem cells, short-lived progenitors, cancer stem cells, or circulating tumour cells, is crucial to acquire a better understanding of normal or diseased tissue biology. To address this challenge we first sequenced the transcriptome of hundreds of randomly selected cells from mouse intestinal organoids<sup>1</sup>, cultured self-organizing epithelial structures that contain all cell lineages of the mammalian intestine. Organoid buds, like intestinal crypts, harbour stem cells that continuously differentiate into a variety of cell types, occurring at widely different abundances<sup>2</sup>. Since available computational methods can only resolve more abundant cell types, we developed RaceID, an algorithm for rare cell type identification in complex populations of single cells. We demonstrate that this algorithm can resolve cell types represented by only a single cell in a population of randomly sampled organoid cells. We use this algorithm to identify *Reg4* as a novel marker for enteroendocrine cells, a rare population of hormone-producing intestinal cells<sup>3</sup>. Next, we use *Reg4* expression to enrich for these rare cells and investigate the heterogeneity within this population. RaceID confirmed the existence of known enteroendocrine lineages, and moreover discovered novel subtypes, which we subsequently validated *in vivo*. Having validated RaceID we then applied the algorithm to *ex vivo*-isolated *Lgr5*-positive stem cells and their direct progeny. We find that *Lgr5*-positive cells represent a homogenous abundant population of stem cells mixed with a rare population of *Lgr5*-positive secretory cells. We envision broad applicability of our method for discovering rare cell types and the corresponding marker genes in healthy and diseased organs.

Single-cell mRNA sequencing has emerged as a powerful method to simultaneously measure cell-to-cell expression variability of thousands of genes<sup>4</sup>. Recently, it was demonstrated that sequencing of randomly selected cells from spleen<sup>5</sup> and lung tissue<sup>6</sup> permits the identification of known cell types within these organs. The approaches used in these and other recently published studies<sup>7–10</sup> show good performance in discovering abundant cell types but cannot detect rare cell types.

To profile cell types of widely varying abundance within a complex mixture we introduce a method for rare cell type identification (RaceID) and apply it to investigate rare cell types in the mouse small intestine<sup>11</sup>.

The continuously self-renewing intestinal epithelium is arranged in crypts and villi. A small number of intestinal stem cells reside near the crypt bottom and give rise to rapidly proliferating transit amplifying (TA) cells. While migrating upward along the crypt–villus axis TA cells

develop into the terminally differentiated cell types<sup>2,12</sup>. Absorptive enterocytes constitute the most abundant cell type, while all other mature cell types contribute only a few percent or less. The secretory lineage comprises mucus producing goblet, hormone secreting enteroendocrine, and Paneth cells, which provide a niche for the stem cell and secrete bactericidal products. In addition, tuft cells are believed to sense the luminal content.

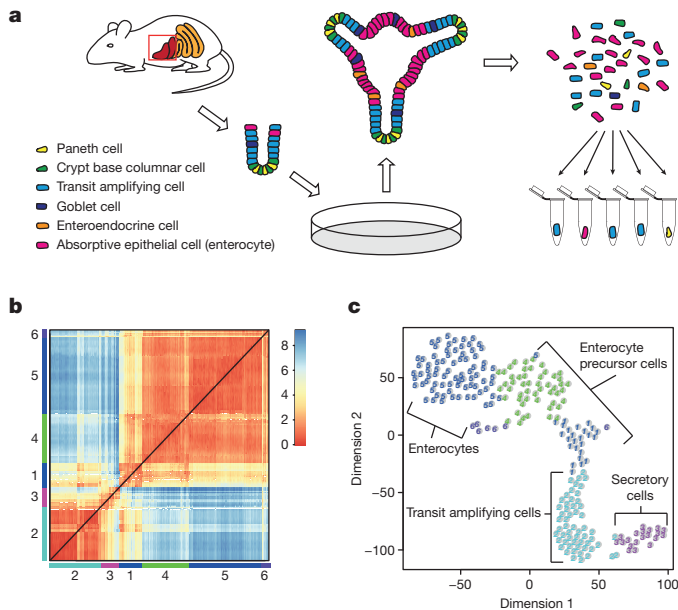
To obtain clean random mixtures of intestinal cells without contamination of non-epithelial cell types, we use intestinal organoids, small epithelial structures containing all major cell types found in the intestinal epithelium<sup>1</sup>. Using a modified version of the cell expression by linear amplification and sequencing (CEL-seq) method<sup>13</sup> incorporating unique molecular identifiers to count transcripts<sup>14</sup> (Fig. 1a and Extended Data Fig. 1), we sequenced 238 randomly selected organoid cells with more than 3,000 transcripts in total each, and quantified 3,777 genes with more than five transcripts in at least one cell.

Hierarchical clustering of the transcriptome correlation matrix suggested the presence of three major groups of cells (Extended Data Fig. 2a). To screen for abundant cell types more systematically we employed *k*-means clustering of the correlation matrix with six clusters as inferred by the gap statistic<sup>15</sup> (see Methods, Fig. 1b and Extended Data Fig. 2b). We visualized these clusters in two dimensions (Fig. 1c) using *t*-distributed stochastic neighbour embedding (t-SNE)<sup>16</sup> and examined if expression of known intestinal marker genes was restricted to specific clusters (Extended Data Fig. 3). The intestinal alkaline phosphatase (*Alpi*) is a known enterocyte marker and showed a gradual expression increase across clusters 1, 4 and 5 (Extended Data Fig. 3a). Cluster 3 comprises distinct cells with non-overlapping expression of marker genes for diverse secretory cell types, such as the enteroendocrine marker *Chga*, the goblet cell marker *Muc2*, or the Paneth cell marker *Lyz1* (Extended Data Fig. 3b–d). The central cluster 2 does not express specific marker genes, but shows pronounced expression of genes encoding ribosomal proteins (Extended Data Fig. 3e), indicating the presence of transit amplifying cells. The bottom part of this cluster contains cells expressing low levels of the stem cell marker *Lgr5* (Extended Data Fig. 3f).

To detect rare cell types, we screened for outliers that could not be explained by a background model accounting for technical and biological gene expression noise (see Methods, Fig. 2a and Extended Data Fig. 4). Distinct outliers were grouped into 10 novel clusters based on transcriptome correlation (see Methods, Fig. 2b). Differential gene expression analysis revealed the presence of rare cell types among these clusters, comprising goblet, tuft, Paneth and enteroendocrine cells (Fig. 2c and Supplementary Table 1). Moreover, RaceID detected three secretory precursor clusters, co-expressing *Neurog3* with *Krt7*, *Pax4* or *Ang4* (Fig. 2c). Available methods for cell type identification<sup>5</sup> were clearly out-performed by RaceID (Extended Data Fig. 5a and Supplementary Note). Extensive experimental validation proved the

<sup>1</sup>Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences), 3584 CT Utrecht, The Netherlands. <sup>2</sup>University Medical Center Utrecht, Cancer Genomics Netherlands, 3584 CG Utrecht, The Netherlands.

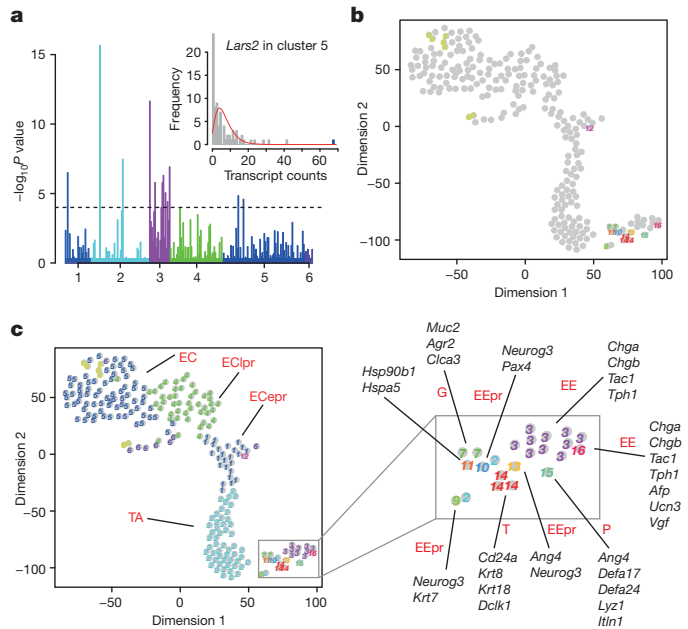
\*These authors contributed equally to this work.



**Figure 1 | Profiling cell composition of mouse intestinal organoids with single-cell sequencing.** **a**, Intestinal crypts were isolated from mice and grown into intestinal organoids as described previously<sup>1</sup>. Organoids were dissociated and single cells, collected by fluorescence-activated cell sorting (FACS), were sequenced by a modified version of the CEL-seq method<sup>13,14</sup> (see Methods). **b**, Heat map indicating similarities between 238 single cells measured by Euclidean distances of the transcriptome correlation matrix (unitless; see Methods). *k*-means clustering identified six major groups of cells colour coded along the axes. **c**, t-SNE map representation of transcriptome similarities between individual cells. Clusters identified in **b** were highlighted with different colours and corresponding intestinal cell types identified on the basis of known marker genes are indicated.

specificity and sensitivity of RaceID (Extended Data Fig. 4c–e and Supplementary Note). We further showed that cell cycle related genes are unlikely to affect the results of RaceID for our data set (Extended Data Fig. 5f).

Enteroendocrine cells control metabolism by secreting at least ten different hormones<sup>3,17</sup> and individual cells produce subsets of those<sup>18,19</sup>. To profile heterogeneity of enteroendocrine cells, we aimed at purifying a random population of mature enteroendocrine cells. We identified enteroendocrine markers by a *z*-score analysis (Fig. 3a). Among the top scoring genes were novel markers such as the proteinase *Pappa2* and the largely uncharacterized gene *Reg4*. We focused on the latter, since it was highly expressed in enteroendocrine cells with hundreds of sequenced mRNAs. We validated *Reg4* as an *in vivo* marker for enteroendocrine cells by single-molecule fluorescent *in situ* hybridization<sup>20</sup> (smFISH) in the mouse intestine. Co-staining of *Chga* and *Reg4* revealed high levels of *Reg4* in enteroendocrine cells and substantially lower levels in Paneth cells at the crypt bottom (Fig. 3b). We then purified *Reg4*-positive organoid cells derived from a *Reg4*-red fluorescent protein (dsRed) reporter mouse (see Methods and Extended Data Fig. 6). RaceID predicted three major groups among the 161 cells surviving our filtering criteria (Fig. 3c, d). Upregulation of defensins suggested that one of these groups comprises maturation stages of the Paneth cell lineage (Extended Data Fig. 6e). Within the second group of cells we identified a contamination with TA cells (Extended Data Fig. 6f, g). The remaining 60 cells (37%) arise from the enteroendocrine lineage proving a pronounced enrichment (~eightfold) of this rare cell. We observed two major subgroups with low and high levels of *Chga*, respectively (Fig. 3d and Extended Data Fig. 6h). Hormones expressed in cells with low levels of *Chga* comprise *Cck*, *Ghrl*, *Sct*, *Nts* and *Gcg* (Extended Data Fig. 7), identifying this group as an intestinal specific branch of enteroendocrine cells<sup>3,18,19</sup>.



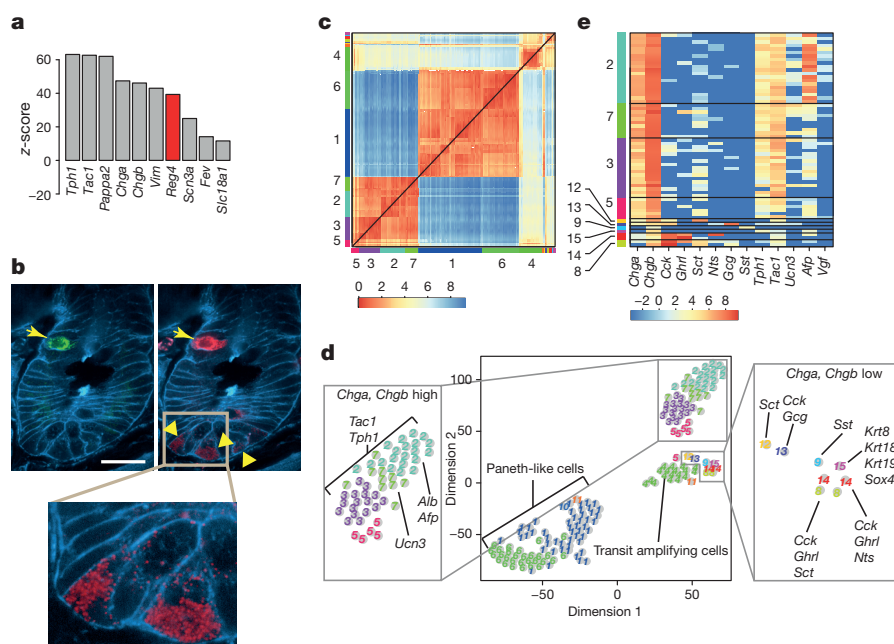
**Figure 2 | RaceID algorithm identifies rare cell types among hundreds of sequenced cells.** **a**, Histogram showing the negative logarithm of the probability that transcript levels in a particular cell are not explained by a background model accounting for the expected variability. Clusters are highlighted as in Fig. 1. The probability threshold for outlier identification ( $10^{-4}$ ) was included (black broken line). The inset explains the derivation of outlier probabilities (see Methods) for the gene *Lars2* in cluster 5. The transcript count histogram (grey) is compared to the background model (red) and an outlier event is highlighted (purple). **b**, t-SNE map with additional clusters highlighted that were inferred from outlier cells (see Methods). **c**, t-SNE map of all clusters obtained by the RaceID algorithm (left) and close-up of all clusters in the secretory lineage (cluster 3 in Fig. 1c). Genes corresponding to known markers specifically upregulated in each cluster of the secretory lineage are indicated. EC, enterocytes; EClpr, late enterocyte precursors; ECepr, early enterocyte precursors; TA, transit amplifying cells; G, goblet cells; EE, enteroendocrine cells; EEpr, enteroendocrine precursors; P, Paneth cells; T, tuft cells.

The other sub-group co-expressed *Tph1* and *Tac1*, indicating hormone production of serotonin and substance P, respectively, and therefore comprises enterochromaffin cells (Fig. 3d and Extended Data Fig. 8a, b).

Within this sub-group RaceID identified three novel subtypes (Fig. 3d, e). In 7 out of 16 (41%) cells within cluster 3 we detected co-expression of *Tac1* and *Cck* (Extended Data Fig. 7a), previously considered to be markers of separate subtypes<sup>3</sup>. Expression of urocortin 3 (*Ucn3*), a ligand of the corticotropin-releasing hormone (*Crh*) receptor type 2 was significantly elevated in cluster 7 ( $P < 4.1 \times 10^{-4}$ , see Methods and Extended Data Fig. 8c). Although colonic expression of *Ucn3* has been described<sup>21</sup>, it was not known to be expressed in the small intestine. Finally, we observed strong upregulation of Albumin (*Alb*,  $P < 4.3 \times 10^{-6}$ , see Methods) and the related alpha-fetoprotein (*Afp*,  $P \sim 0$ , see Methods) in cluster 2 (Extended Data Fig. 8d). Albumin can bind lipophilic hormones and could regulate their accessibility<sup>22</sup>. In the same cluster, we measured upregulation of VGF nerve growth factor (Extended Data Fig. 8e). See Extended Data Fig. 8f and Supplementary Table 2 for additional marker genes.

We validated the existence of the novel enterochromaffin subtypes *in vivo* at the mRNA and protein level by conducting smFISH and immunofluorescence experiments in the mouse intestinal epithelium (Fig. 4, Extended Data Fig. 9, Methods and Supplementary Table 4).

After having shown that RaceID can discriminate cell types we wanted to test our method on stem cells marked by *Lgr5* expression. Heterogeneity of the intestinal stem cell pool is still controversial<sup>23–27</sup> and single-cell sequencing could help to better characterize this



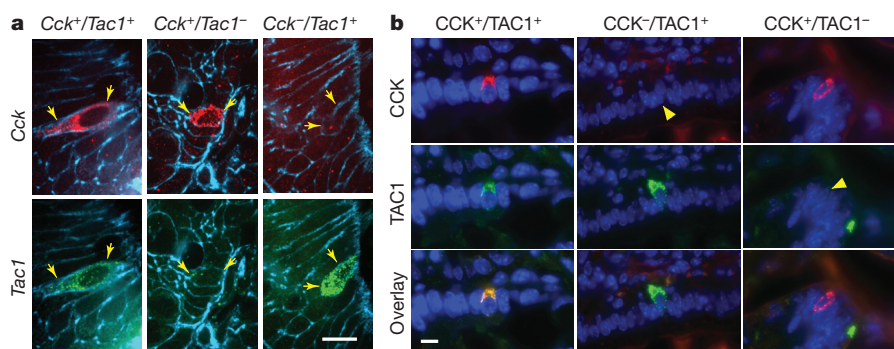
**Figure 3 | *Reg4* is a novel marker of differentiated enteroendocrine cells.** **a**, Histogram of top ten z-scores for upregulation in differentiated enteroendocrine cells. For each gene, we extracted the average level observed in mature enteroendocrine cells (cluster 3 and 16), subtracted the average level observed across all remaining cells, and divided by the standard deviation of transcript levels in these cells. **b**, Validation of *Reg4* as an *in vivo* marker of mature enteroendocrine cells by single-molecule fluorescent *in situ* hybridization (smFISH). Cryosections of mouse small intestine were hybridized with smFISH probes against *Chga*, conjugated to tetramethylrhodamine (TMR, green) and *Reg4*, conjugated to cyanine 5 (Cy5, red). Cell borders were visualized with AlexaFluor 488-conjugated phalloidin (blue). The enteroendocrine cell (arrow) expresses a high level of *Chga* and co-expresses *Reg4*.

Paneth cells at the crypt bottom (arrowheads and inset) express strongly reduced levels of *Reg4*. Scale bar, 20  $\mu$ m. **c**, Heat map representing the transcriptome similarities measured by the Euclidean distance of the transcriptome correlation matrix (see Methods) for *Reg4*-positive cells. RaceID clusters are colour-coded along the axes. Cluster numbers are shown for the bigger clusters. **d**, t-SNE map showing all clusters identified by RaceID for *Reg4*-positive cells. Different colours and numbers highlight distinct clusters. Selected upregulated genes are shown for individual clusters ( $P < 10^{-3}$ , see Methods). Close-ups are shown for clusters of enteroendocrine cells. **e**, Heat map of hormone expression ( $\log_2$  scale) in subtypes of enteroendocrine cells identified by RaceID. The two groups of cells with low and high expression of *Chga*, respectively, display distinct patterns.

population. We first used intestinal organoids derived from an Lgr5–green fluorescent protein (GFP) reporter mouse<sup>23</sup> to purify and sequence 96 Lgr5–GFP<sup>+</sup> cells. RaceID detected only a single large cluster and few outliers that were mostly Paneth cells (Extended Data Fig. 10), suggesting that intestinal stem cells represent a uniform population. Since a distinct reserve pool of quiescent Lgr5-positive cells has been suggested<sup>25–27</sup>, we next tried to characterize a population of *ex vivo* isolated Lgr5 expressing cells. For this experiment,

we isolated 192 Lgr5–enhanced GFP<sup>+</sup> (EGFP<sup>+</sup>) cells from the small intestine of an Lgr5–EGFP reporter mouse<sup>24</sup> (Supplementary Table 5). RaceID classified these cells into a single large homogenous cluster and a few outliers (Fig. 5a, b).

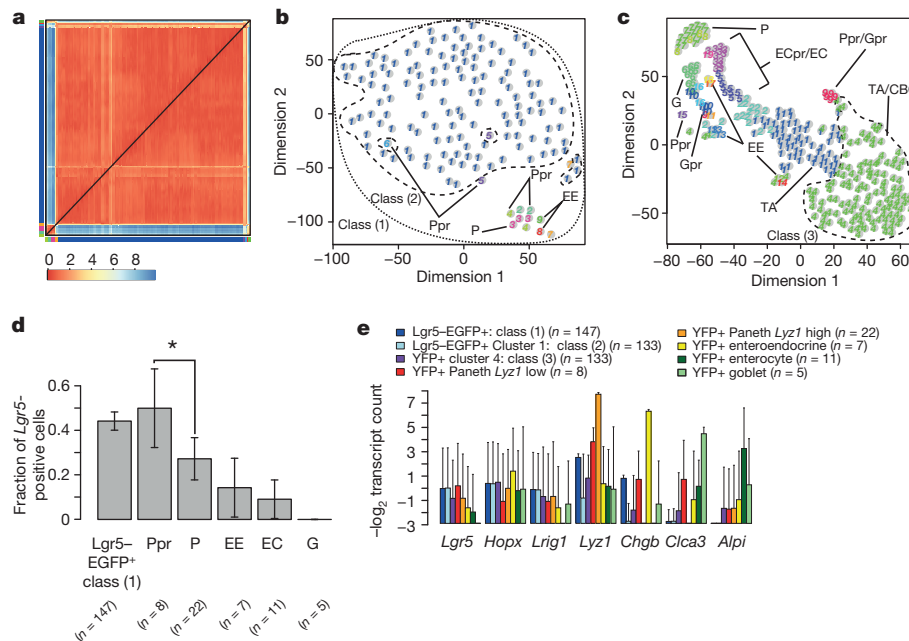
As a complementary approach, we traced the progeny of Lgr5-positive cells *in vivo*, using a reporter mouse that expresses CreERT2 from an Lgr5 promoter and YFP from a Rosa26 promoter with a loxP-flanked transcriptional roadblock. Administration of tamoxifen leads



**Figure 4 | Single-molecule FISH and immunofluorescence experiments confirm expression of markers for enteroendocrine cell sub-populations in the mouse small intestine.** **a**, Small intestine cryosections were hybridized with smFISH probe libraries. Scale bar, 10  $\mu$ m. *Cck* and *Tac1* are expressed by a subset of enteroendocrine cells. Probes against *Cck*, conjugated to Cy5 (upper panel, red) and against *Tac1*, conjugated to TMR (lower panel, green) were used for hybridization. Cell borders were visualized by staining with phalloidin, conjugated with AlexaFluor 488 (blue). Enteroendocrine cells co-expressing the two markers were observed (left), as well as cells expressing only *Cck*

(middle) or *Tac1* (right). Arrows point at cell borders. **b**, Immunostaining was performed on cryosections of mouse small intestinal tissue. Scale bar, 20  $\mu$ m. Expression of CCK and TAC1 was observed in a subset of enteroendocrine cells. CCK and TAC1 were visualized by indirect immunostaining with antibodies against CCK (upper panel, red) and against TAC1 (middle panel, green). Nuclei were counterstained with DAPI (blue). Rare cells, co-expressing the two markers, were observed (left), as well as cells, expressing only TAC1 (middle), or only CCK (right). Arrowheads point at CCK or TAC1-negative cells.





**Figure 5 | Characterization of *ex vivo* isolated *Lgr5*-EGFP<sup>+</sup> cells.** **a**, Heat map of transcriptome similarities measured by the Euclidean distance of the transcriptome correlation matrix (see Methods) for *Lgr5*-EGFP<sup>+</sup> cells purified using an *Lgr5*-EGFP reporter mouse<sup>24</sup>. RaceID clusters are colour coded along the axes. Colours correspond to panel **b**. **b**, t-SNE map of RaceID clusters for *Lgr5*-EGFP<sup>+</sup> cells. Cells of cluster 7 express non-coding RNAs (*Malat1*, *Kcnq1ot1*) and could not be characterized. The other cell types were assigned based on marker gene expression. The stem cell classes (1) (dotted line) and (2) (dashed line) are outlined (see text for details). **c**, t-SNE map of RaceID clusters for 5-day lineage tracing progeny of *Lgr5*-positive (YFP<sup>+</sup>) cells. Cell types identified based on marker genes are indicated. The stem cell class (3) is outlined (dashed line, see text for details). **d**, Fraction of *Lgr5*-positive cells for

the major intestinal lineages. See main text for details. Error bars and *P* value were inferred from binomial statistics reflecting uncertainty due to sampling. The number of cells (*n*) is indicated below the plot. \**P* < 0.05. **e**, Mean expression of marker genes in different sets of *Lgr5*-positive cells and in cells of the major intestinal lineages. Class (1) corresponds to all *Lgr5*-EGFP<sup>+</sup> cells shown in **b**, while class (2) and (3) correspond to the sub-populations indicated by the dashed line in **b** and **c**, respectively. Error bars indicate the standard deviation across cells. The number of cells (*n*) is given in the legend. EC, enterocytes; ECpr, enterocyte precursors; TA, transit amplifying cells; CBC, crypt base columnar cells; G, goblet cells; Gpr, goblet cell precursors; EE, enteroendocrine cells; P, Paneth cells; Ppr, Paneth cell precursors; Ppr/Gpr, early precursors of the Paneth and goblet cell lineage.

to yellow fluorescence protein (YFP) protein production in *Lgr5*-positive cells and their progeny. We sequenced 432 YFP-positive cells collected five days after label induction (Supplementary Table 5). As expected, RaceID detected differentiated cells of all major lineages, with a relatively large proportion of Paneth cells (Fig. 5c). A possible explanation for the over-representation of Paneth cells is label induction in mature Paneth cells or their precursors<sup>27</sup>. We then quantified the fraction of *Lgr5*-positive cells in all major lineages. To extract cells of a lineage independent of the maturation state, we only required >5 transcripts of a lineage marker (*Lyz1* for Paneth cells, *Chgb* for enteroendocrine cells, *Alpi* for enterocytes, and *Clca3* for goblet cells). Paneth cells were split into early and late stages, with *Lyz1* expression lower or higher than the median, respectively. While we detected *Lgr5* transcripts in ~45% of the *Lgr5*-EGFP<sup>+</sup> cells, this fraction was lower than 15% for most of the other major cell types (Fig. 5d). Only for Paneth cells we observed an elevated proportion of *Lgr5*-positive cells, which was significantly higher in early (~50%) compared to late Paneth cells (~28%) (Fig. 5d). This could be due to the *Lgr5* RNA half-life exceeding the rapid transition time of stem cells into Paneth cells, leading to a transient state where stem and Paneth cell genes are co-expressed.

To examine if the population of *Lgr5* expressing cells show any kind of fate bias towards a particular lineage, we first distinguished three classes of stem cells: (1) all *Lgr5*-EGFP<sup>+</sup> cells (Fig. 5b), (2) the subset of *Lgr5*-EGFP<sup>+</sup> cells after removal of the outliers identified by RaceID (Cluster 1, Fig. 5b), and (3) the stem cell/early TA cluster from the lineage tracing data (Cluster 4, Fig. 5c). Based on the RaceID prediction we consider class (2) as a homogenous pool of stem cells, while class (1) contains a few additional *Lgr5* expressing cells of other lineages. Class (3) represents the homogenous stem cell population identified by RaceID in the lineage tracing data and is thus expected to resemble

class (2). We then performed a marker gene analysis in all three classes and, for comparison, in cell populations of all major lineages (Fig. 5e). The three classes of stem cells showed similar expression of *Lgr5* and other stem cell markers (*Hopx* and *Lrig1*). Expression of secretory lineage markers (*Lyz1* and *Chgb*) was substantially lower in class (2) and (3) compared to class (1), and did not exceed the background level observed in any other lineage (Fig. 5e). This argues against additional secretory cells in class (2) and (3) that could have remained undetected by RaceID. Elevated expression of *Lyz1* and *Chgb* in class (1) is thus solely due to the few *Lgr5*-EGFP<sup>+</sup> secretory cells identified by RaceID (Fig. 5b). Interestingly, *Lgr5* transcript levels in early Paneth cells were similar to those in stem cells and reduced in the other cell types (Fig. 5e).

Taken together, we conclude that, both in organoids and *in vivo*, *Lgr5*-positive cells represent a homogenous population of cells mixed with a rare population of Paneth and enteroendocrine cells. In comparison to the other lineages, Paneth cells express the highest level of *Lgr5*, consistent with the observation that Paneth cell precursors can revert to the stem-cell state upon tissue damage<sup>25,27</sup>. It remains a possibility that high *Lgr5* expression in Paneth cells is an artefact due to sequencing doublets of Paneth and crypt bottom stem cells. However, we consider this unlikely, because expression of *Lgr5* is significantly elevated in early versus late Paneth cells (Fig. 5d). Finally, we would like to caution that heterogeneity among lowly expressed genes could still exist within the stem cell pool, which would be invisible owing to the limited sensitivity of current single cell sequencing protocols. Alternatively, stem cell heterogeneity could extend to *Lgr5*-low cells as described previously<sup>26</sup>, which are not captured by our enrichment strategy.

In summary, we demonstrated here the ability of RaceID to correctly classify different cell types in a complex mixture and reveal heterogeneity

among rare cells. We believe that single-cell mRNA sequencing in combination with the RaceID algorithm is a powerful tool to unravel heterogeneity of rare cell types in both healthy and diseased organs.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 14 November 2014; accepted 23 July 2015.**

**Published online 19 August 2015.**

- Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71**, 241–260 (2009).
- Engelstoft, M. S., Egerod, K. L., Lund, M. L. & Schwartz, T. W. Enteroendocrine cell types revisited. *Curr. Opin. Pharmacol.* **13**, 912–921 (2013).
- Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Rev. Genet.* **14**, 618–630 (2013).
- Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
- Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
- Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
- Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
- Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature Biotechnol.* **32**, 1053–1058 (2014).
- Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
- Clevers, H. The intestinal crypt, a prototype stem cell compartment. *Cell* **154**, 274–284 (2013).
- Barker, N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nature Rev. Mol. Cell Biol.* **15**, 19–33 (2014).
- Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
- Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nature Methods* **11**, 637–640 (2014).
- Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. B.* **63**, 411–423 (2001).
- Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2570–2605 (2008).
- Schonhoff, S. E., Giel-Moloney, M. & Leiter, A. B. Minireview: Development and differentiation of gut endocrine cells. *Endocrinology* **145**, 2639–2644 (2004).
- Habib, A. M. *et al.* Overlap of endocrine hormone expression in the mouse intestine revealed by transcriptional profiling and flow cytometry. *Endocrinology* **153**, 3054–3065 (2012).
- Egerod, K. L. *et al.* A major lineage of enteroendocrine cells coexpress CCK, secretin, GLP, GLP-1, PYY, and neurotensin but not somatostatin. *Endocrinology* **153**, 5782–5795 (2012).
- Raj, A., Van Den Bogaard, P., Rifkin, S. A., Van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5**, 877–879 (2008).
- Saruta, M. *et al.* Urocortin 3/stresscopin in human colon: possible modulators of gastrointestinal function during stressful conditions. *Peptides* **26**, 1196–1206 (2005).
- Baker, M. E. Albumin, steroid hormones and the origin of vertebrates. *J. Endocrinol.* **175**, 121–127 (2002).
- Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
- Roth, S. *et al.* Paneth cells in intestinal homeostasis and tissue injury. *PLoS ONE* **7**, e38965 (2012).
- Li, N. *et al.* Single-cell analysis of proxy reporter allele-marked epithelial cells establishes intestinal stem cell hierarchy. *Stem Cell Rep.* **3**, 876–891 (2014).
- Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495**, 65–69 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by an European Research Council Advanced grant (ERC-AdG 294325-GeneNoiseControl) and a Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) Vici award.

**Author Contributions** D.G., A.L. and A.v.O. conceived the project. D.G. developed the algorithm, performed all computations and wrote the manuscript. A.L., L.K. and K.W. performed all sequencing experiments. A.L. performed the lineage tracing experiment and all imaging experiments. N.S. made the *Reg4*-dsRed mouse and was supervised by H.C.; O.B. contributed the Lgr5-GFP organoids. A.L., L.K., N.S. and H.C. edited the manuscript. A.v.O. guided experiments, data analysis and writing of the manuscript, and edited the manuscript.

**Author Information** RNA-seq data are deposited in Gene Expression Omnibus, accession number GSE62270. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.v.O. ([a.vanoudenaarden@hubrecht.eu](mailto:a.vanoudenaarden@hubrecht.eu)).

## METHODS

No statistical methods were used to predetermine sample size, and the experiments were not randomized.

**Generation of the *Reg4*-dsRed mouse.** *Reg4*-dsRed knock-in mice were generated by homologous recombination in embryonic stem cells by targeting a diphtheria toxin receptor-2A peptide-dsRed express2 cassette to the ATG start codon of *Reg4*. Generation of the knock-in mouse and experiments were performed according to guidelines and reviewed by the Dier Experimenten Commissie (DEC) of the KNAW.

***Lgr5*-GFP organoids.** Organoids from *Lgr5*-GFP-DTR reporter mice<sup>23</sup> were cultured under standard conditions<sup>1</sup>. Nine days after splitting, the organoids were dissociated into single cells using TrypLE (Invitrogen) for 15 min at 37 °C and mechanical disruption using a glass Pasteur pipette. Cells were washed twice in advanced DMEM/F12 (GIBCO) and resuspended in advanced DMEM/F12 + 4 µg µl<sup>-1</sup> DNase I (Roche) and Propidium Iodide (Sigma). Among PI negative fraction, high level of GFP (top 10%) expressing cells were sorted directly in 96-wells containing 100 µl TRIzol.

**Lineage tracing experiments.** For lineage tracing experiments we injected 0.4 mg tamoxifen into 3-month-old *Lgr5*-CreERT2 C57Bl6/J mice crossed to a Rosa26LSL-YFP reporter mice.

**Isolation of crypts from mouse small intestine.** Crypts were isolated from *Reg4*-dsRed mouse as described previously<sup>1</sup>. Briefly, the whole of the small intestine was dissected, flushed with cold Ca<sup>2+</sup>- and Mg<sup>2+</sup>-free PBS and cut to 4–5 cm pieces for convenience. Intestines were cut open longitudinally and villi were scraped off with a glass slide. Intestine fragments were washed twice with cold Ca<sup>2+</sup>- and Mg<sup>2+</sup>-free PBS, then incubated with 5 mM EDTA in PBS at 4 °C for 30 min, with gentle agitation. Crypts were released by vigorous shaking of the tissue fragments, pelleted by centrifugation (200g at 4 °C for 5 min), washed once with cold PBS and once with Advanced DMEM/F12 medium (Life Technologies), pelleted by centrifugation and used to generate organoids.

**Isolation of *Lgr5*-EGFP<sup>+</sup> cells from the mouse intestine.** Freshly isolated small intestines of *Lgr5*-EGFP-IRES-creERT2 mice<sup>24</sup> were incised along their length and villi were removed by gentle scraping. The tissue was then washed in ice-cold PBS0 and subsequently incubated in PBS0/EDTA (5 mM) for 5 min followed by gentle shaking to remove remaining villi. The intestine was then incubated in PBS/EDTA for 30 min at 4 °C. Vigorous shaking yielded free crypts that were incubated in MEM (Gibco) supplemented with trypsin (2 mg ml<sup>-1</sup>; Sigma) and DNase I (2,000 U ml<sup>-1</sup>; Sigma) for 30 min at 37 °C. Subsequently, cells were spun down, resuspended in MEM/DNase and DAPI (Life Technologies) and filtered through a 40-µm mesh. DAPI-negative, GFP-expressing cells were directly sorted in TRIzol (Life Technologies) using a BD FACSAria II cell sorter (BD Bioscience).

**Intestinal organoid culture.** Villin-Cre organoids were a gift from H. Farin. Organoid culture was carried out as described<sup>1</sup>. Briefly, organoids were grown in a drop of Matrigel (BD Biosciences), overlaid with the ENR medium (see below). Organoids were passaged weekly, with 1:4 dilution. Briefly, the old medium was aspirated, old Matrigel drop broken, organoids were washed and pelleted by centrifugation at 200–300g, then mixed with fresh Matrigel. Drops of Matrigel-organoid mix were placed on the bottom of tissue culture dish, let to solidify, and overlaid with ERM culture medium. *Reg4*-dsRed organoids were derived from the *Reg4*-dsRed mouse as described<sup>1</sup>. Briefly, small intestinal crypts were isolated, pelleted and mixed with Matrigel. Crypt-Matrigel mix was plated and cultured in the ENR medium. Newly generated organoids were expanded in culture for at least three weeks before harvesting for the experiment. The ENR medium is an Advanced DMEM/F12 medium, supplemented with penicillin/streptomycin, 1× GlutaMAX, 10 mmol l<sup>-1</sup> HEPES, 1×B27 (Life Technologies), Noggin (conditioned medium, 10% volume), R-spondin 1 (conditioned medium, 10% volume), 1 mmol l<sup>-1</sup> N-acetylcysteine (Sigma) and 50 ng ml<sup>-1</sup> recombinant mouse EGF (Peprotech). The R-spol 1 and Noggin conditioned media were generated in HEK293T cells, stably expressing HA-mouse Rspo1-Fc (gift from Calvin Kuo, Stanford University), or transiently transfected with mouse Noggin-Fc plasmid. Advanced DMEM/F12 with penicillin/streptomycin, 10 mmol l<sup>-1</sup> HEPES, and 1× GlutaMAX was conditioned for 1 week.

**Single-cell suspension preparation, FACS sorting and RNA extraction.** Organoids were dissociated to single cells as previously<sup>28</sup>, with a few modifications. Briefly, medium was removed from organoid cultures, organoids were resuspended in TrypLE, incubated at 37 °C for 5–15 min, with passaging through a glass pipette each 5 min and microscopic monitoring. Upon are disruption of most of the cell aggregates, cells were pelleted by centrifugation (5 min at 300–400g), washed twice with Advanced DMEM/F12 with 10% fetal calf serum, resuspended in the same medium. Cells were strained through 20 µm mesh filter and stained with either propidium iodide (Villin-Cre cells), or DAPI (*Reg4*-dsRed cells). Single cells were then sorted by flow cytometry (FACS Aria, BD). Each cell was

sorted directly into single wells of 96-well PCR plates, each well containing 100 µl TRIzol reagent. Identical quantity of the ERCC Spike-in RNA (0.03 µl of 1:50,000 dilution) was added to each well. Total RNA was extracted from each cell, according to the TRIzol manufacturer's protocol with a few alterations. To facilitate RNA precipitation, 0.2 µl of GlycoBlue reagent (Life Technologies) was added to each sample. Isopropanol precipitation was carried out overnight. RNA pellets were air dried for up to 15 min, then resuspended in the CEL-Seq first-strand primer solution.

**Control library with mouse embryonic stem cells and fibroblasts.** Irradiated mouse embryonic fibroblasts were cultured in DMEM containing 10% FBS (Gibco), 2 mM GlutaMAX (Gibco), 0.1 mM MEM nonessential amino acids and 1% Pen/Strep (Gibco). Wild-type mouse embryonic stem cells were derived from C57BL6 mice and cultured in DMEM containing 10% FBS (Gibco), 2 mM GlutaMAX (Gibco), 0.1 mM MEM nonessential amino acids, 1% Pen/Strep (Gibco) and 1,000 U LIF ml<sup>-1</sup> (ESGRO). The control library contained 5 ESCs with barcodes 1–5, 5 MEFs with barcodes 6–10, 75 mouse small intestinal organoids cells with barcodes 11–85, 6 controls without template but with reverse transcription primer, barcodes 87–92 and 5 empty controls.

**Pool-and-split control sample preparation.** Organoids, derived from the *Reg4*-dsRed knock-in mouse, were dissociated to a single-cell suspension and dsRed-positive cells were sorted using FACS, as described above. Pools of 100 cells were collected into single tubes containing 100 µl TRIzol reagent and processed for the total RNA extraction. Total RNA from 100 cells was resuspended in nuclease-free water and ERCC Spike-in RNA was added to the RNA solution (3 µl of 1:50,000 dilution, per 100 pooled cells). Then each of the pooled samples was split into 100 separate portions. CEL-Seq was performed on 200 resulting split samples.

**Tissue preparation and immunofluorescence.** Freshly dissected mouse small intestines were flushed and fixed in cold 4% paraformaldehyde in PBS for 3 h. After fixation, the intestines were incubated in cryoprotective solution (30% sucrose in PBS) at 4 °C overnight, then frozen blocks were prepared in Tissue-Tek OC compound (VWR) and stored at –80 °C. Five µm thick cryosections were cut and mounted on poly-L-lysine (Sigma)-coated cover glass. Sections were fixed in 4% paraformaldehyde for 15 min, permeabilized with 0.25% Triton X100 (Sigma) on PBS for 5 min and blocked for 1 h in PBS containing 0.2% Triton X100, 1% BSA (Sigma), and 2% each normal donkey (Jackson) and goat (Monosan) serum. Next, the sections were incubated for 1 h with primary antibody solution, (in PBS with 1% BSA), washed, and incubated for 1 h with secondary antibody (in PBS). Nuclei counterstain was done with DAPI (100 ng ml<sup>-1</sup> in PBS) for 10 min. The sections were then washed and mounted in Fluoromount-G (Electron Microscopy Sciences). Imaging was done on Leica fluorescence microscope with a 100× oil immersion objective, using MetaMorph imaging software. Images were processed and combined using ImageJ and Photoshop programs.

**Antibodies.** For indirect immunofluorescence the following antibodies were used: anti-mouse Urocortin 3 rabbit polyclonal antibody (Yanaihara Institute, Y364); anti-mouse CCK rabbit polyclonal antibody (LifeSpan BioSciences, aa26-33, LS-C190673); anti-mouse VGF rabbit polyclonal antibody (Abcam, ab69899); anti-mouse Tac1 guinea pig polyclonal antibody (Abcam, ab10353). The following secondary antibodies were used: Goat anti-rabbit IgG, Cy5-conjugated (Life Technologies, A10523); Donkey-anti guinea pig IgG, TRITC-conjugated (Jackson Labs, 706-025-148). For direct immunofluorescence we used the mouse monoclonal antibody against AFP, conjugated to Alexa Fluor 594 (Cell Signaling, 7877).

**CEL-seq library preparation.** Single cells were processed using the previously described CEL-seq technique<sup>13</sup>, with several modifications. A 4-bp random barcode as unique molecular identifier (UMI) was added to the primer in between the cell specific barcode and the poly T stretch (Supplementary Table 3). Dried RNA, prepared from single cells by TRIzol extraction method, was resuspended in primer solution, denatured at 70 °C for 2 min and quickly chilled, after which the first strand synthesis mix was added. The rest of the protocol was carried out as published<sup>13</sup>, with no substantial alterations. Libraries were sequenced on an Illumina HighSeq 2500 using 50 bp paired end sequencing.

**Single-molecule FISH.** Probe libraries were designed and fluorescently labelled as previously described<sup>20</sup>. All probe libraries consist of 20 to 39 oligonucleotides of 20-bp length (see Supplementary Table 3 for probe sequences) complementary to the coding sequence of the genes. Cells were hybridized overnight with probes at 30 °C, as previously described<sup>20</sup>. DAPI and phalloidin-AlexaFluor488 staining was done after washes. Images were acquired on a Perkin-Elmer Spinning Disc confocal microscope with a 100× oil-immersion objective (numerical aperture 1.4) using Perkin Elmer Velocity software. Images were recorded as stacks with a z spacing of 0.3 µm. Diffraction-limited dots corresponding to single mRNA molecules were automatically detected using custom Matlab software, based on previously described algorithms<sup>20</sup>. Briefly, the images were first filtered using a three-dimensional Laplacian of Gaussian filter, followed by selection of the intensity



threshold at which the number of connected components was least sensitive to the threshold.

**Quantification of transcript abundance.** Paired end reads obtained by CEL-seq were aligned to the transcriptome using bwa<sup>29</sup> (version 0.6.2-r126) with default parameters. The transcriptome contained all RefSeq gene models based on the mouse genome release mm10 downloaded from the UCSC genome browser<sup>30</sup> and contained 31,109 isoforms derived from 23,480 gene loci. All isoforms of the same gene were merged to a single gene locus. The right mate of each read pair was mapped to the ensemble of all gene loci and to the set of 92 ERCC spike-ins<sup>31</sup> in sense direction. Reads mapping to multiple loci were discarded. The left read contains the barcode information: the first eight bases correspond to the cell specific barcode followed by 4 bases representing the unique molecular identifier (Supplementary Table 3). The remainder of the left read contains a polyT stretch followed by few (<15) transcript-derived bases. The left read was not used for quantification. For each cell barcode we counted the number of unique molecular identifiers for every transcript and aggregated this number across all transcripts derived from the same gene locus. Based on binomial statistics we converted the number of observed unique molecular identifiers into transcript counts<sup>14</sup>.

**Rare cell type identification algorithm RaceID.** *Data preparation.* The clustering algorithm takes as input a matrix with transcript counts for all genes in each cell. As a first preprocessing step cells with low overall transcript counts are removed. We require at least 3,000 transcripts per cell for the whole organoid data and 1,000 transcripts per cell for the *Reg4*-positive cells. For the latter we observed overall lower transcript numbers. Next, the total transcript count within each cell is normalized to the median transcript number across cells. Alternatively, downsampling of the transcript pool to the required minimal total transcript count can be applied. Hereafter, we add a pseudocount of 0.1 to the expression value of each gene to avoid divergences when computing fold changes. In the next step lowly expressed genes are filtered out. Genes that are not expressed with a minimum of five transcripts for the whole organoids data and three transcripts for the *Reg4*-positive cells in at least a single cell are discarded. Furthermore, highly expressed genes that saturate the pool of available UMIs (>500 transcripts after normalization for the whole organoid data and >2,000 transcripts after normalization for the *Reg4*-positive cells) are discarded, since these genes potentially introduce artefacts in the clustering. For the *Reg4*-positive cells we amended this last filtering step, since hormones crucial for the cell type determination saturated the UMIs in only very few cells. We tested the robustness of the RaceID predictions using the more relaxed setting also for the random organoid cells (Extended Data Fig. 5e) and the more stringent settings for the *Reg4*-positive cells (Extended Data Fig. 6d), respectively. In each case, we observed the same rare cell types for the cells that survive the filtering criteria of both settings. We also analysed *Lgr5*-positive intestinal cells. Here we applied the same filtering criteria as used for the whole organoid data, since the single cell sequencing yielded a high number of transcripts per cell. For all *ex vivo* isolated cells we also applied these settings. However, the data were downsampled to the same transcript number in all cells (that is, from all cells subsets of transcripts are sampled with a size corresponding to the minimal total transcript count across all cells surviving the filtering step). This approach was applied since batch effects due to combining different libraries were more pronounced and downsampling reduces technical noise caused by variation in library complexity.

**k-means clustering.** The clustering step of RaceID identifies larger clusters of different cells by *k*-means clustering. First, a similarity matrix is computed that contains Pearson's correlation coefficients for all pairs of cells. Subtracting the coefficients from one yields a distance matrix, which serves as input for *k*-means clustering. *k*-means clustering is applied to the similarity matrix using the Euclidean metric. In comparison to direct clustering of the expression matrix this approach yielded improved cluster separation. The number of clusters used for *k*-means clustering is determined from the gap statistic<sup>15</sup>, that is, from the difference of the average within cluster dispersion in uniformly distributed and in the actual data. By default, the cluster number is determined as the first local maximum of the gap statistic where the maximum exceeds its neighbours by >25% of their standard deviation. If the gap statistic does not exhibit a clear maximum, the cluster number demarcating the point where the gap statistic starts to saturate should be used as input for the *k*-means clustering. Given the number of clusters, *k*-means clustering of the distance matrix is performed and cluster reproducibility is assessed by bootstrapping using the clusterboot function of the R package fpc. The algorithm computes Jaccard's similarity to quantify cluster reproducibility. If more than a single cluster has a Jaccard's similarity lower than 0.5, the clustering

should be repeated with fewer clusters. Importantly, the outlier identification step of the algorithm will correct for an underestimation of the actual cluster number and it is thus recommended to start with a conservative estimate.

**Identification of outlier cells.** To identify outlier cells within each cluster the algorithm evaluates transcript count variability of every gene across all cells in this cluster. The expected baseline level of expression variability, quantified by the transcript count variance, is inferred from the ensemble of all cells. A second order polynomial is fitted to the transcript count variance as a function of the average transcript count in logarithmic space (Extended Data Fig. 4a). In comparison to a linear regression the polynomial leads to a significant improvement of the regression as was tested by ANOVA model comparison for the data sets presented here (ANOVA  $P < 2.2 \times 10^{-16}$  for both data sets). For the random organoid cells the residual sum of squares was reduced by 20% when using the second order polynomial instead of the linear regression. This polynomial serves as an estimate of the expected variance-mean dependence under the assumption, that the majority of genes do not exhibit cluster (or cell type) specific expression. Next, each cluster is screened for outlier cells by computing the transcript count probability in each cell for a given gene from a negative binomial distribution defined by the average transcript count of this gene across all cells in the cluster and the expected variance computed by the second order polynomial. Assuming a lower limit of Poissonian noise, values of the expected variance lower than the mean are replaced by the mean. In practice, this does not happen and the regression yields noise estimates well above the Poissonian limit for all data sets analysed so far. If the multiple testing (Benjamini-Hochberg) corrected transcript count probability of a specified number of genes (two for our data) is below a defined probability threshold ( $<10^{-4}$  for our data) in a given cell, this cell is considered an outlier. The total number of outliers can be plotted as a function of the probability threshold (Extended Data Fig. 4b), which should be chosen such that it separates the tail of this distribution from the bulk behaviour.

**Inference of final clusters.** Given the set of outlier cells the final clusters that should largely correspond to different cell-types or -states are inferred. To this end, outlier cells are first merged to outlier clusters if their transcriptome correlation exceeds the 75%-quantile of the distribution of cell-to-cell correlation within the original clusters after outlier removal. Subsequently, new cluster centres are computed for the remaining original and the new outlier clusters by averaging transcript counts within these clusters and each cell is reassigned to the most highly correlated cluster centre.

**Two-dimensional representation of cell type maps.** For visual inspection of cell clusters and associated cell types we apply a dimensional reduction of cell-to-cell distances as computed by the distance matrix (see above) using a machine learning algorithm termed *t*-distributed stochastic neighbour embedding (*t*-SNE)<sup>16</sup>. Briefly, this algorithm converts the original point-to-point distance distribution to a lower dimensional Student's *t*-distribution. The location of all points in the map is determined by a stochastic minimization of the Kullback-Leibler divergence of the original distances with respect to the mapped distances.

**Identification of differentially expressed genes.** To identify genes that were on average up- or downregulated within a cluster compared to the ensemble of all cells, the fold change in absolute transcript counts was computed after normalizing the total transcript count in a cell to the median transcript count within the cluster under consideration. As shown previously<sup>14</sup> for single cell sequencing data, median normalized transcript levels exhibit Poissonian noise for most genes with small deviations at high expression. A *P* value for significant up- or downregulation was therefore computed based on Poissonian statistics and multiple testing corrected by the Benjamini-Hochberg method.

**Code availability.** The RaceID algorithm is supplied as an R script (Supplementary Data 1) along with sample code (Supplementary Data 2), an extensive reference manual (Supplementary Data 3) and sample data (Supplementary Table 6), corresponding to the random organoid cell transcriptome data analysed in this paper. Bug fixes and updates of RaceID can be downloaded from <https://github.com/dgrun/RaceID>.

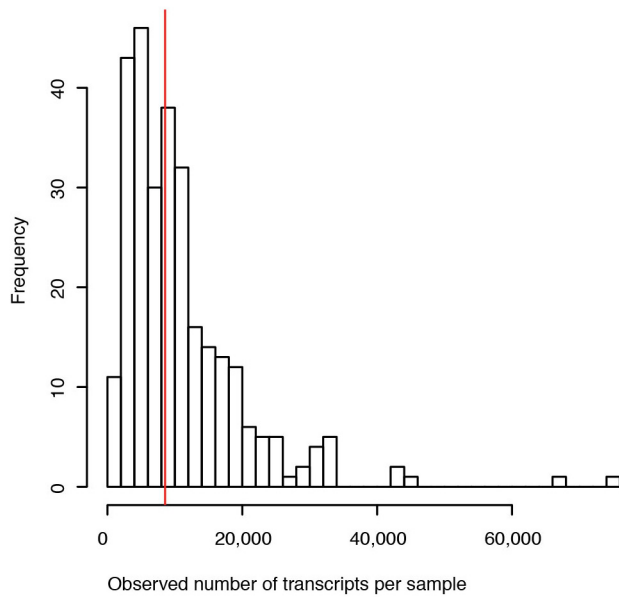
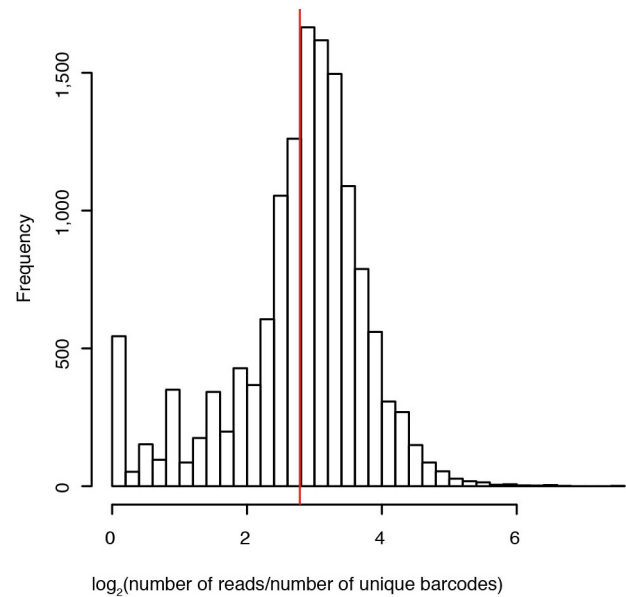
28. Yin, X. *et al.* Niche-independent high-purity cultures of *Lgr5*<sup>+</sup> intestinal stem cells and their progeny. *Nature Methods* **11**, 106–112 (2014).

29. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

30. Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.* **41**, D64–D69 (2013).

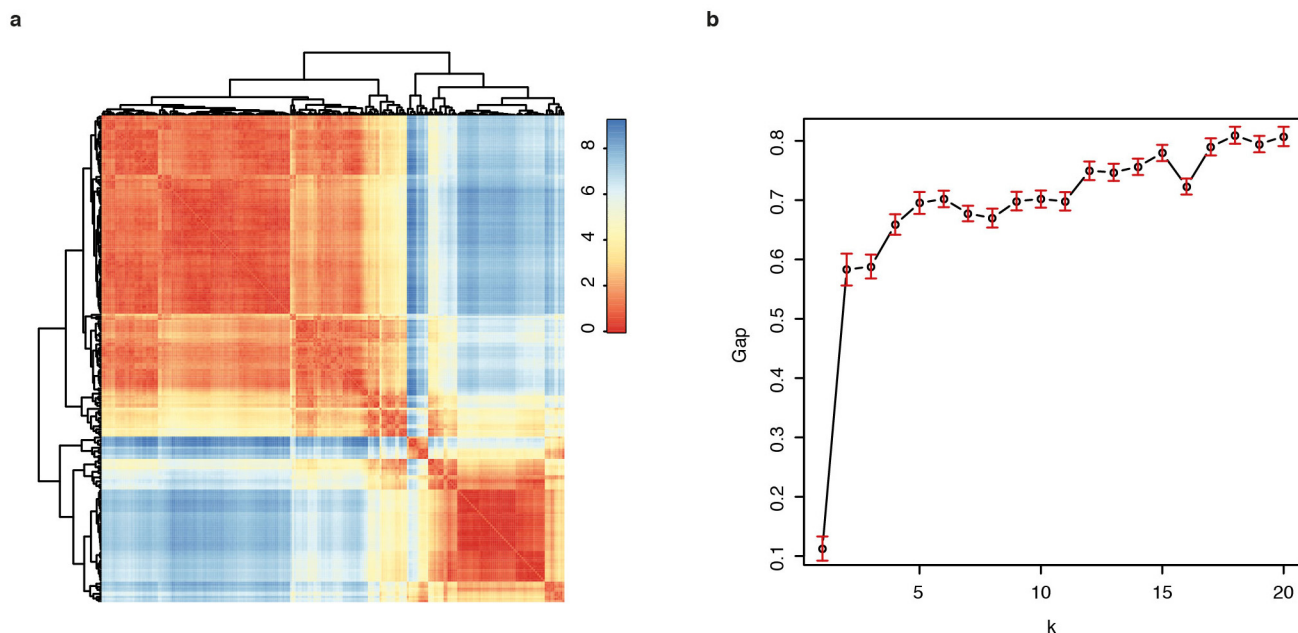
31. The External RNA Controls Consortium. a progress report. *Nature Methods* **2**, 731–734 (2005).



**a****b**

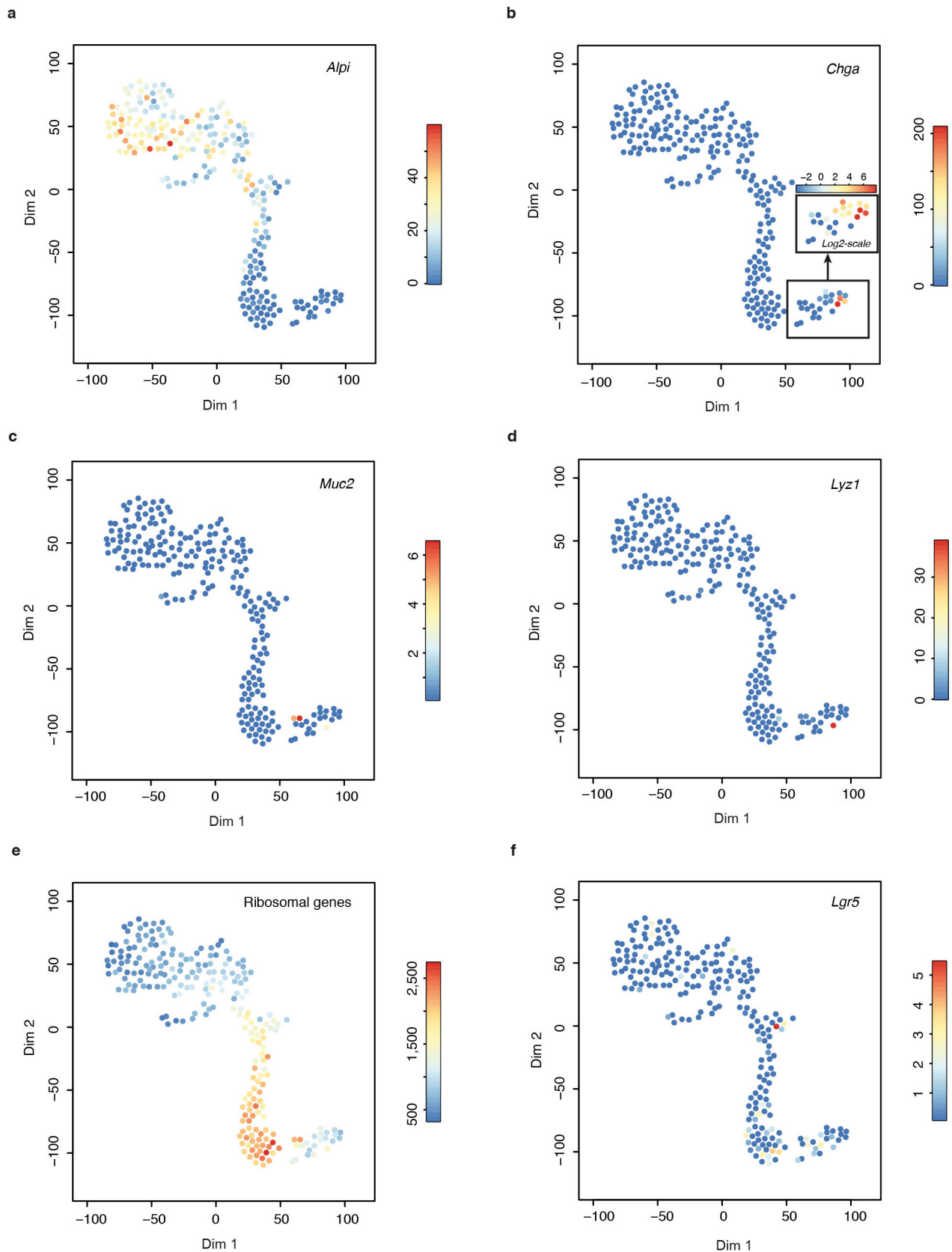
**Extended Data Figure 1 | Sequencing statistics.** **a**, Histogram of the number of sequenced transcripts per cell. The median (red line) is 8,559. The 288 cells were sequenced on two lanes to a total depth of 106,950,038 reads. Of those, 32,694,069 (31%) were mapped with a valid cell-specific barcode. **b**, Histogram

of the total number of reads per cell divided by the total number of sequenced transcripts as counted with unique molecular identifiers. The average level of oversequencing across all genes is 6.9-fold (red line).



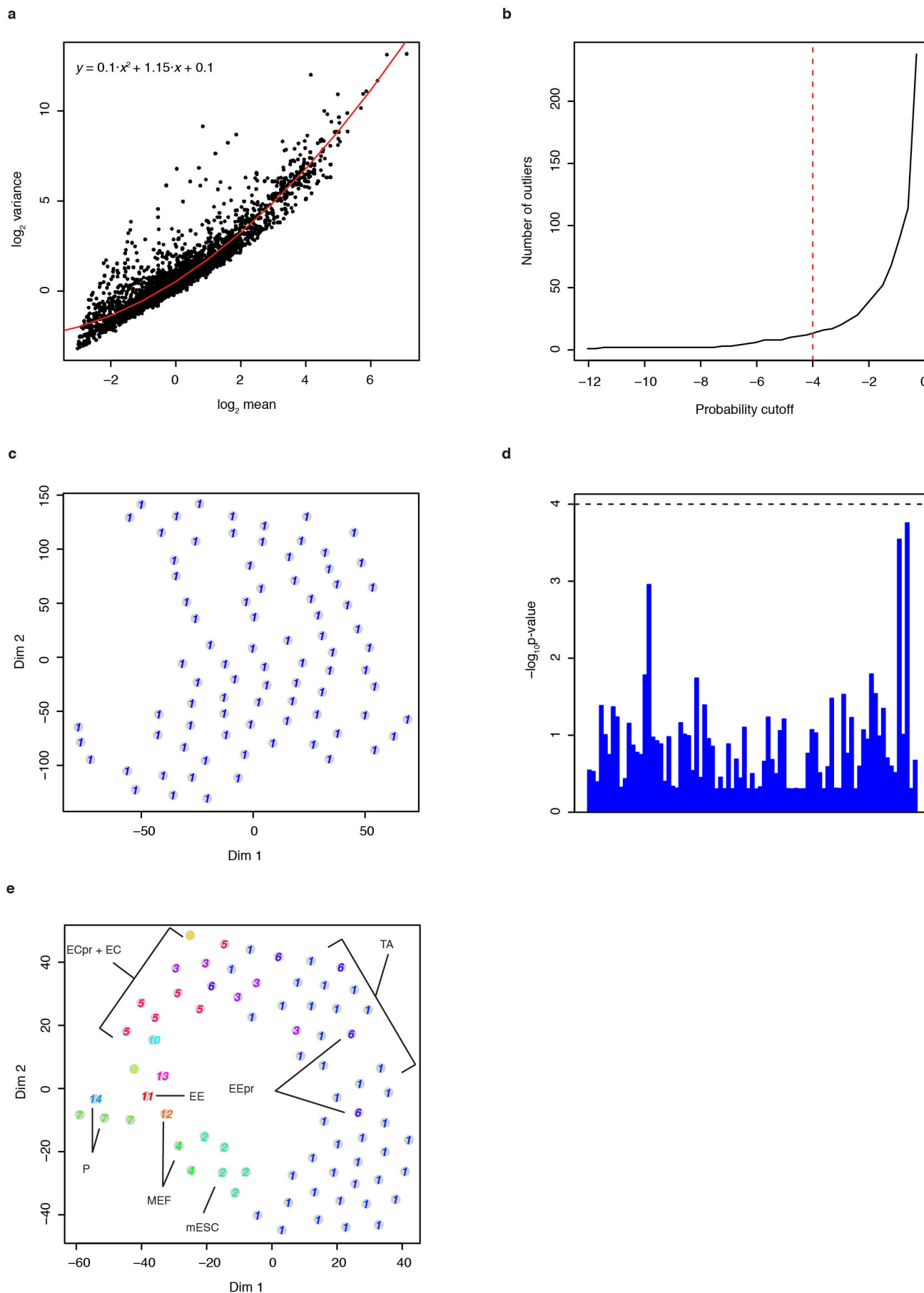
**Extended Data Figure 2 | Clustering reveals major transcriptome differences between intestinal cells.** **a**, Dendrogram obtained by hierarchical clustering of the transcriptome correlation matrix of 238 intestinal cells that survived all filtering steps using Euclidean distance metric. At least three distinct groups of cells can be recognized. **b**, Gap statistic of  $k$ -means clustering of the correlation matrix as a function of the cluster number. The gap statistic

reflects the difference of the average within cluster distance of points in uniformly distributed data and the actual data. The first local maximum provides a good estimate for the number of clusters that achieves optimal separation of the data into clusters<sup>15</sup>. Data points and error bars represent mean and standard deviation across 50 bootstrap samples. For the intestinal cells a number of six clusters was predicted on the basis of the gap statistic.



**Extended Data Figure 3 | Marker gene expression reveals intestinal cell types.** a–f, In the t-SNE maps the transcript count of known marker genes is colour-coded. Shown are maps for the enterocyte specific gene *Alpi* (a), the enteroendocrine marker *Chga* (b), the goblet cell marker *Muc2* (c), the Paneth cell marker *Lyz1* (d), for transcript counts aggregated across all ribosomal genes

(e) and for the stem cell marker *Lgr5* (f). The latter two are upregulated in cells of cluster 2, for which no other specific markers could be identified. These cells most likely correspond to transit-amplifying cells. The inset shown in b depicts transcript counts of *Chga* on a logarithmic scale, since the dynamic range of this gene was very large. Dim, dimension.



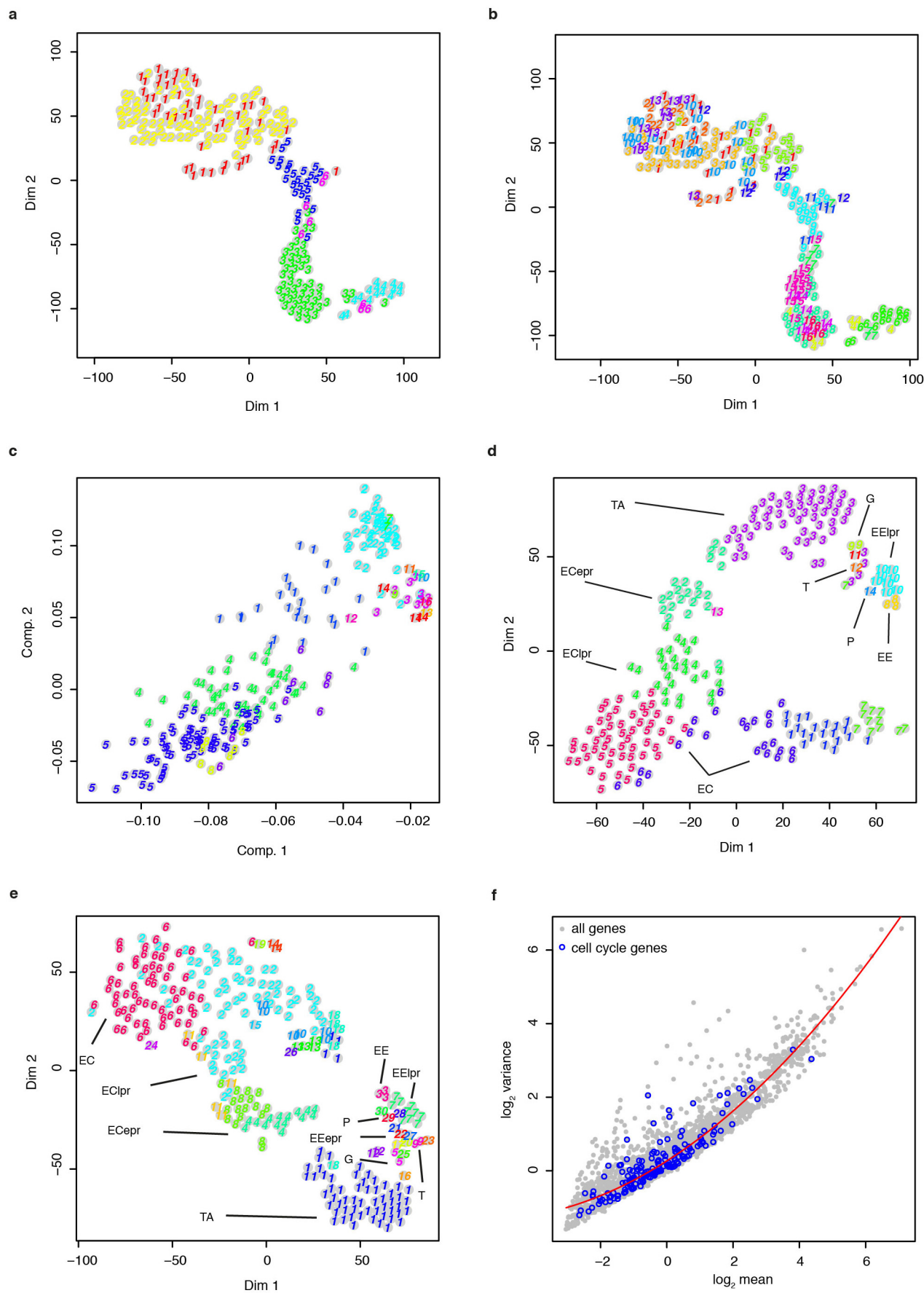


**Extended Data Figure 4 | Identification of rare cell types with RaceID.**

**a**, Variance of transcript count across the entire ensemble of sequenced intestinal cells as a function of mean expression. The red line represents a second order polynomial (upper left corner) that was fitted to the data. Assuming that a large number of genes follows a similar transcript count distribution across different cell types this function can be used to estimate the parameters for a negative binomial that represents a background model for the expected transcript count variability at a given mean expression. The probability of observing a given transcript count in a particular cell of a cluster can be computed using this distribution with the average expression within this cluster as input. If the expression of at least two genes has a probability  $<10^{-4}$  after multiple testing correction a cell is considered an outlier.

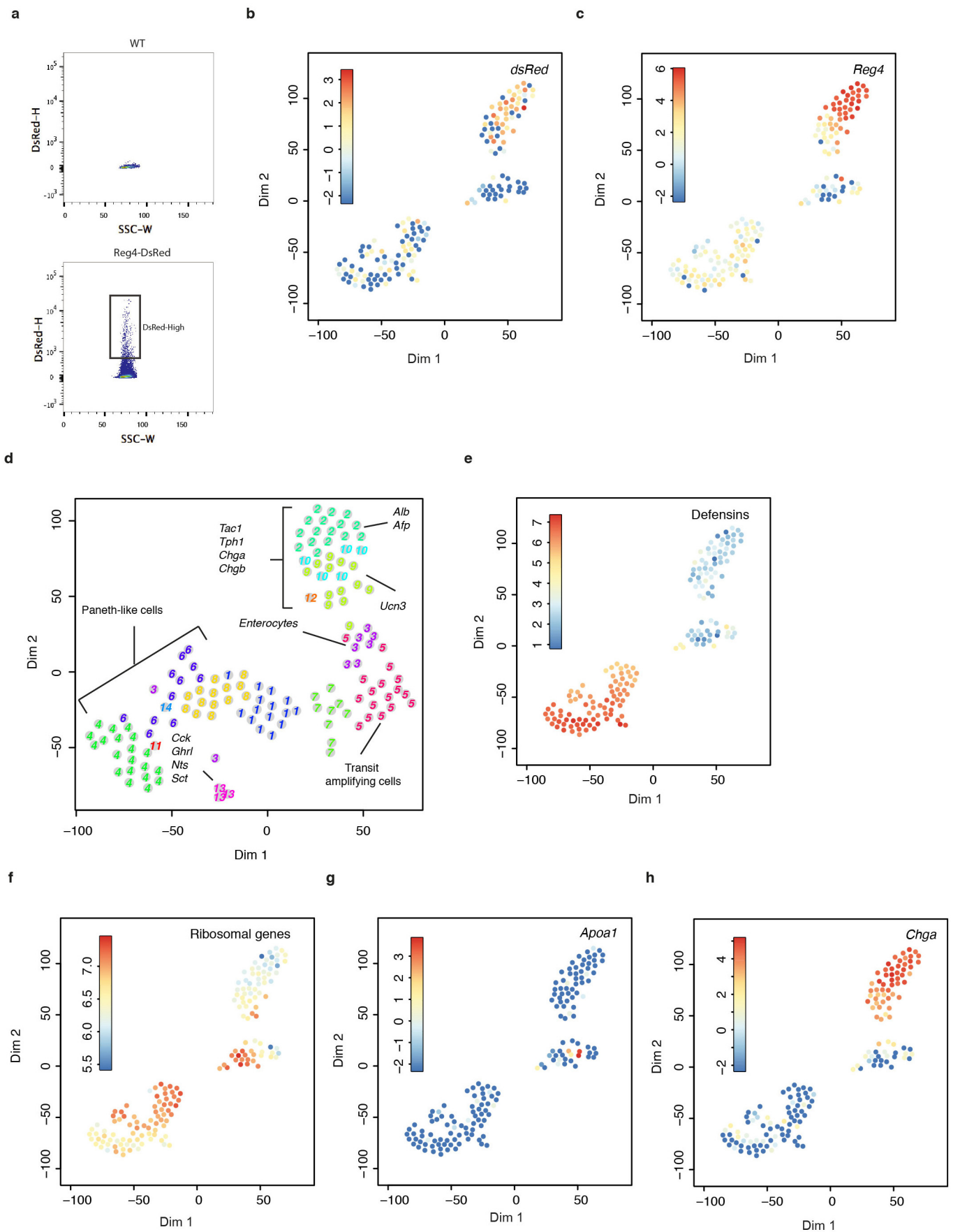
**b**, Number of outlier cells as a function of the probability threshold. The

threshold used in this study ( $10^{-4}$ ) is indicated (red broken line). **c**, **d**, RaceID of pool-and-split controls reflects a low false-positive rate (see Supplementary Note). **c**, t-SNE map of 93 pool-and-split controls. RaceID identifies only a single large cluster with no outliers. **d**, Outlier probability for all pool-and-split controls. The p-value of all cells is higher than the default threshold for outlier identification ( $10^{-4}$ ). **e**, RaceID on a defined mixture of cells demonstrates high specificity (see Supplementary Note). RaceID clusters for a mixture of cells comprising 75 random organoid cells, 5 mouse embryonic stem cells (mESCs) and 5 mouse embryonic fibroblasts (MEFs). Two out of five MEFs did not pass the filtering criteria ( $>3,000$  transcripts per cell). EC, enterocytes; ECpr, enterocyte precursors; TA, transit amplifying cells; EE, enteroendocrine cells; EEpr, enteroendocrine precursors; P, Paneth cells. Dim, dimension.



**Extended Data Figure 5 | Benchmarking RaceID.** **a, b,** To benchmark the RaceID algorithm we compared to a previously published method developed to distinguish cell types from the spleen<sup>5</sup> in thousands of sequenced single cells. We implemented the method according to the specification provided in the original paper. A shortcoming of the method is that it has to be initialized with an expected number  $K$  of cell types. Running the algorithm with  $K = 6$  (**a**) yields results very similar to  $k$ -means clustering with  $K = 6$  (Fig. 1c). However, when the algorithm is run with a larger number of cell types, e. g.  $K = 20$  (**b**), rare cells of the secretory lineage can still not be separated while clusters corresponding to relatively uniform cell types fall apart. We conclude that this algorithm performs well for more abundant cell types but cannot identify rare cell types. **c,** Principal component analysis (PCA) of the transcriptome similarities. The cell types identified by RaceID are highlighted. The first two principal components can only classify major groups of enterocytes, transit amplifying cells, and secretory cells. **d,** To demonstrate that our method is not affected by technical noise due to varying detection efficiency across individual cells, we downsampled the transcriptome of all cells with  $>3,000$  transcripts to the same size, given by the minimum total transcript

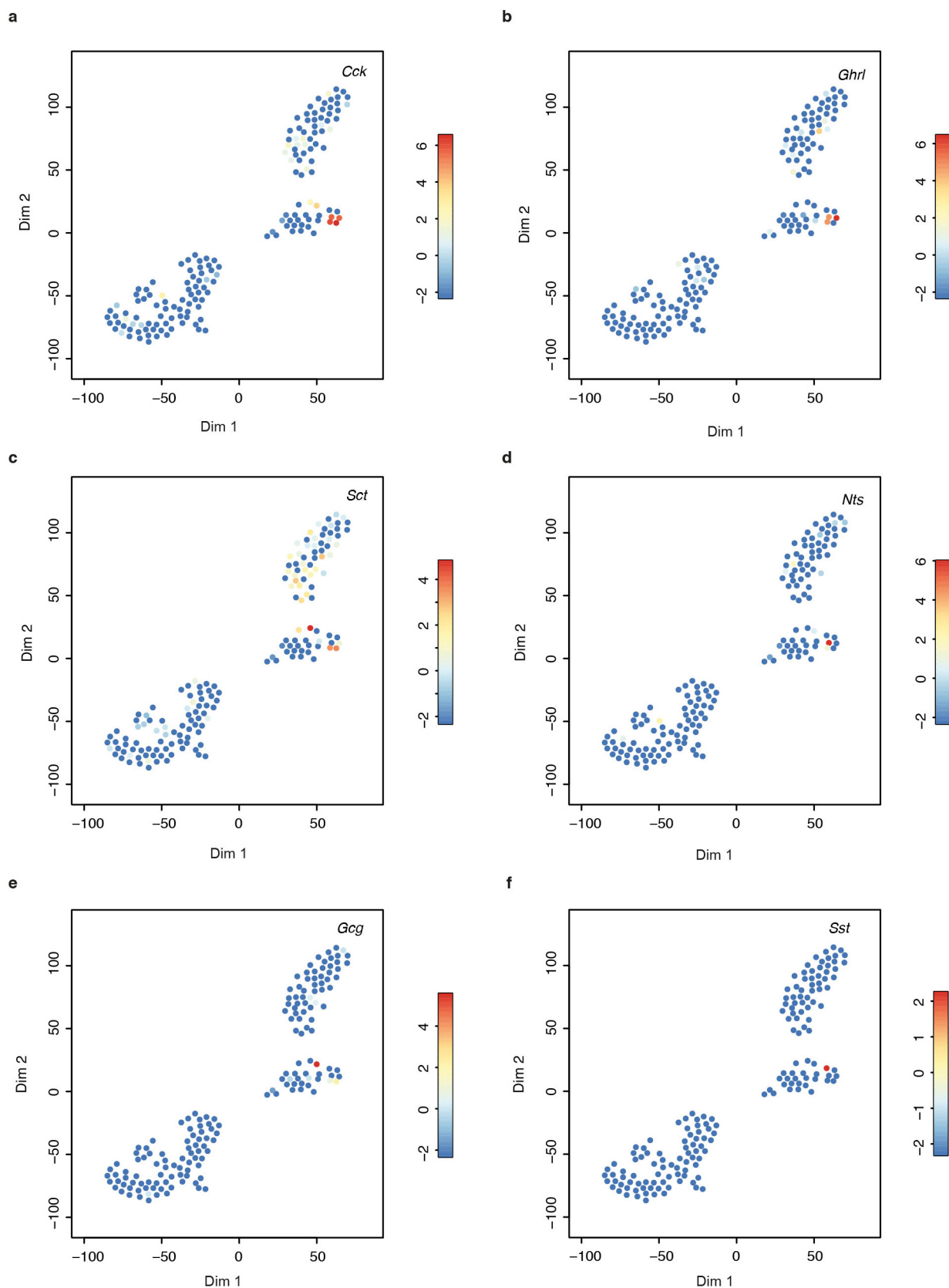
counts across all cells that passed the filtering, and repeated the outlier identification. The t-SNE map shows all cell types identified by this strategy and the results are highly consistent with the normalization-based approach. **e,** The t-SNE map shows the results of RaceID run with relaxed filtering constraints ( $>1,000$  transcripts per cell and only genes with more than three transcripts in at least one cell) as used for the *Reg4*-positive organoid cells. All the rare secretory cell types identified with the original settings were recovered. The different stages of enterocyte differentiation are also apparent. EC, enterocytes; EClpr, late enterocyte precursors; ECepr, early enterocyte precursors; TA, transit amplifying cells; G, goblet cells; EE, enteroendocrine cells; EEIpr, late enteroendocrine precursors; EEepr, early enteroendocrine precursors; P, Paneth cells; T, tuft cells. **f,** Same as Extended Data Fig. 4a, but cell cycle related genes are highlighted as blue circles. This set of genes comprises all genes containing “cell cycle” within their associated “biological process” Gene Ontology terms. Cell cycle related genes do not show increased variability and are thus unlikely to lead to false positives in the outlier detection by RaceID. Dim, dimension.





**Extended Data Figure 6 | Purification of *Reg4*-positive cells from reporter mouse organoids.** In total, 288 cells were sequenced. Ninety-six and 192 cells were analysed from two separate sorting experiments. **a**, Single small intestine cells derived from the wild-type (WT, upper panel) and *Reg4*-dsRed (lower panel) mice were sorted by FACS. Live cells were gated as DsRed-positive (lower panel, gate denoted by black rectangle, DsRed-High). SSC-W, side scatter width, DsRed-H, DsRed height. A median number of 2,813 transcripts per cell were quantified. After filtering out cells with <1,000 transcripts and genes with <3 transcripts in all cells or >2,000 transcripts in a single cells, 161 cells remained for analysis. **b**, **c**, In the t-SNE maps of *Reg4*-positive cells the transcript count of dsRed, driven by a *Reg4* promoter (**b**) and endogenous *Reg4* (**c**) are colour-coded on a logarithmic ( $\log_2$ ) scale. Assuming a previously estimated sensitivity of our sequencing protocol<sup>14</sup>, we measure ~10% of all expressed transcripts. Reporter expression is about eightfold reduced in comparison to endogenous *Reg4*, but expression of both the reporter gene and *Reg4* is strongest in the *Tac1/Tph1* expressing enteroendocrine cells, while expression in Paneth-like cells is reduced. Noticeably, expression of *Reg4*

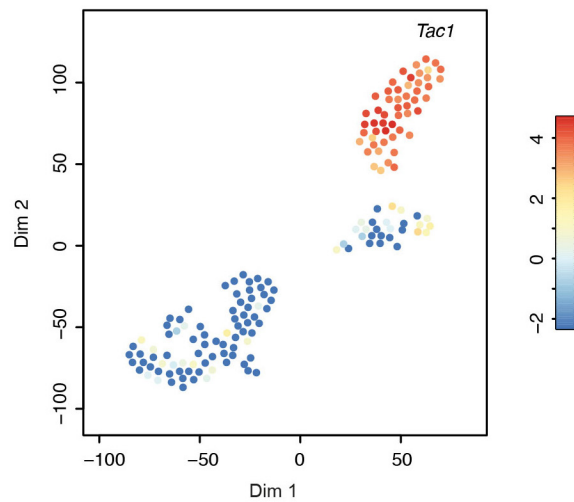
and the reporter gene is also reduced in the *Cck*-positive enteroendocrine cells, similar to *Chga*. **d**, The t-SNE map shows the results of RaceID with more stringent filtering constraints (>3,000 transcripts per cell and only genes with a minimum of five transcripts in at least one cell) as used for the random organoid cells. Although the total number of cells is reduced to 135, most subtypes and rare cells identified with the relaxed settings are still observed, including the *Afp* and *Alb* expressing sub-types, the *Ucn3*-positive cells, the *Cck*-positive cells, the contamination by enterocytes and transit amplifying cells as well as the different Paneth cell states. **e–h**, Marker gene expression reveals intestinal cell types among *Reg4*-positive cells. In the t-SNE maps of *Reg4*-positive cells the transcript count of known marker genes is colour-coded on a logarithmic ( $\log_2$ ) scale. Shown are maps, for transcript counts aggregated across all defensin genes which are highly expressed in Paneth cells (**e**), for transcript counts aggregated across all ribosomal genes (**f**), for the enterocyte marker *Apoa1* (**g**) and for the enteroendocrine markers *Chga* (**h**). Dim, dimension.



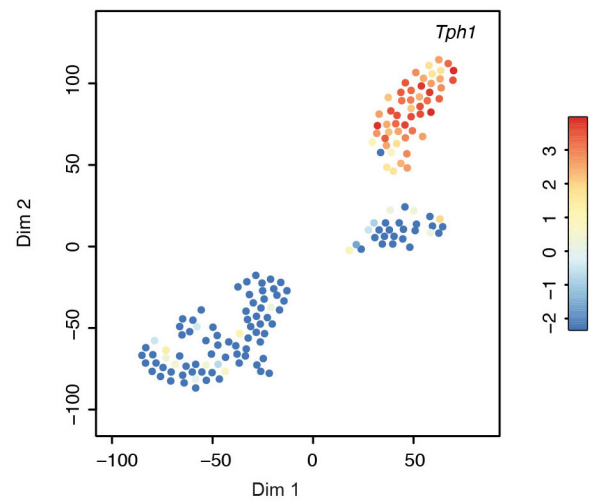
**Extended Data Figure 7 | Expression of intestinal hormones in *Reg4*-positive cells.** In the t-SNE maps of *Reg4*-positive cells the transcript count of hormone encoding genes is colour-coded on a logarithmic ( $\log_2$ ) scale.

a–f, Shown are maps for cholecystokinin (*Cck*) (a), ghrelin (*Ghrl*) (b), secretin (*Sct*) (c), neurotensin (*Nts*) (d), proglucagon (*Gcg*) (e), and somatostatin (*Sst*) (f). Dim, dimension.

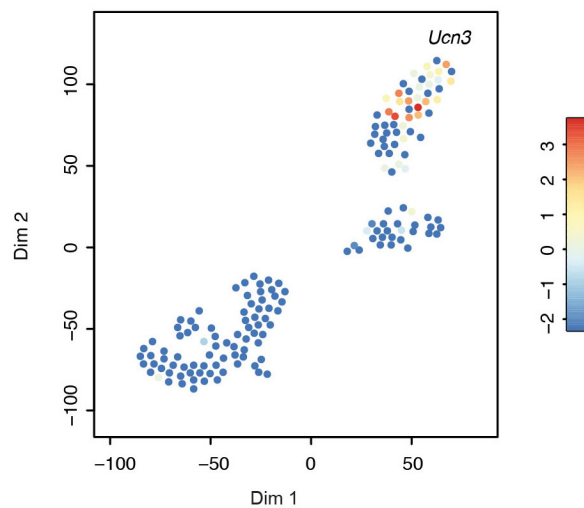
a



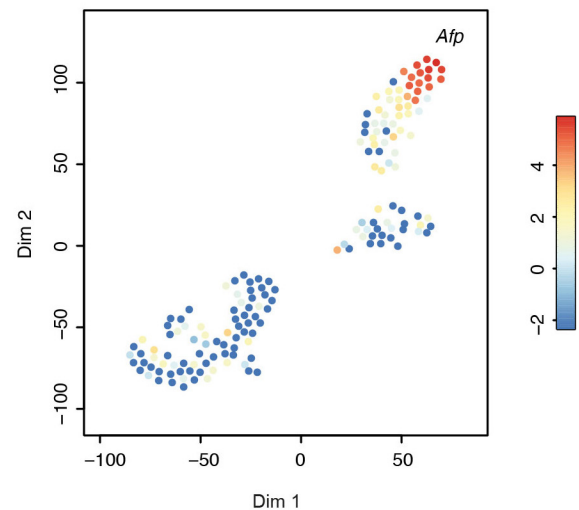
b



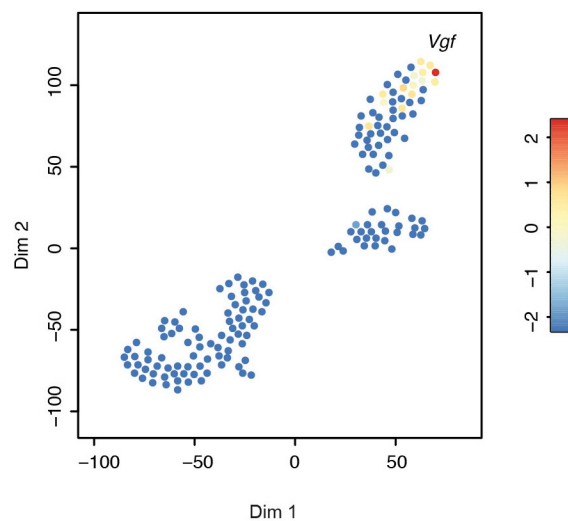
c



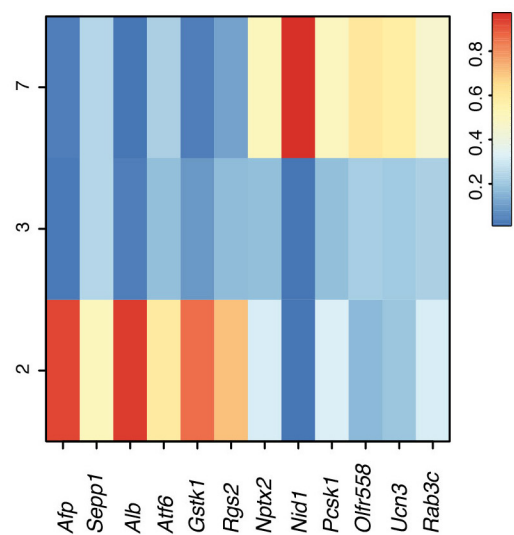
d



e



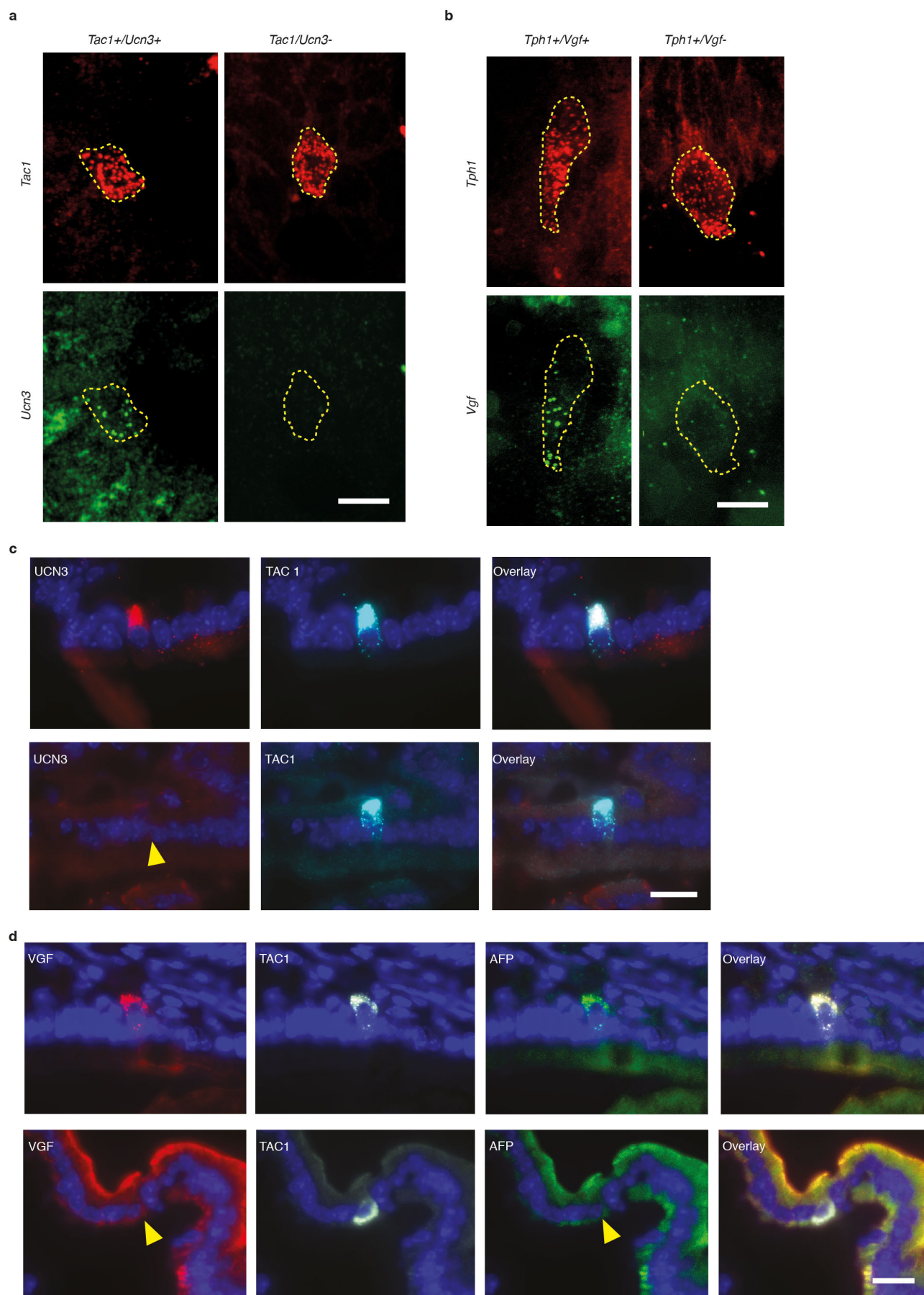
f



**Extended Data Figure 8 | Heterogeneity of substance P producing enteroendocrine cells.** In the t-SNE maps of *Reg4*-positive cells the transcript count of genes specifically expressed in subtypes of enteroendocrine cells is colour-coded on a logarithmic ( $\log_2$ ) scale. Shown are maps for tachykinin (*Tac1*), which encodes substance P, (a), tryptophan hydroxylase 1 (*Tph1*) (b), urocortin 3 (*Ucn3*) (c), alpha-fetoprotein (*Afp*) (d), and VGF nerve growth factor inducible (*Vgf*) (e). f, The heat map shows the average expression of inferred marker genes for the enterochromaffin subtypes (cluster 2, 3 and 7).

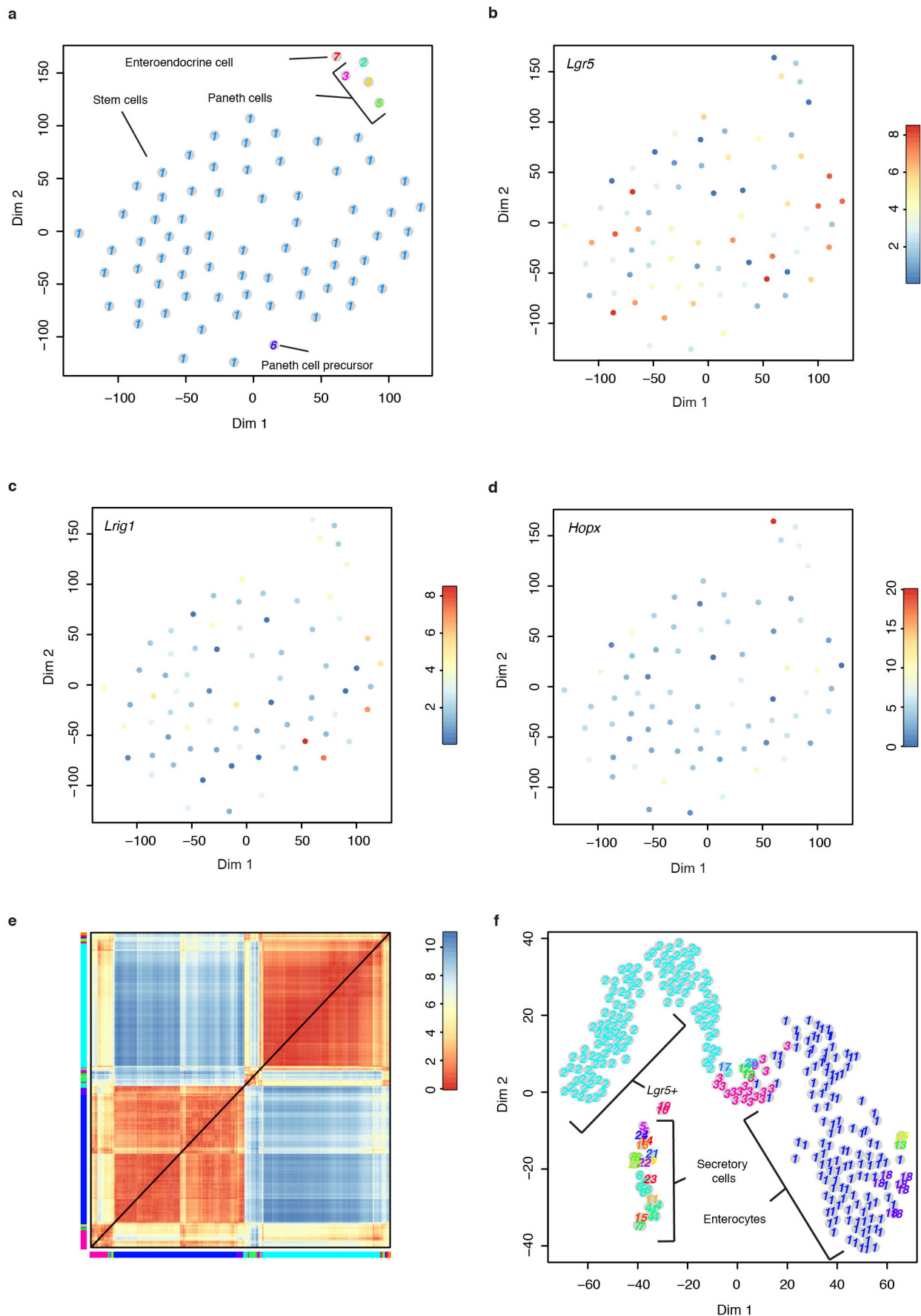
To show all genes on the same scale, the sum of average expression levels in each of the three clusters was normalized to one. Expression levels are shown for six cluster 2 markers and for six cluster 7 markers. Cluster 3 is distinct by the downregulation of both sets. Cluster 5 (not shown here) does not have specific markers and differs from the other clusters by lower expression of mature enterochromaffin markers (*Chga*, *Chgb*, *Tac1*, *Tph1*). This cluster likely comprises non-mature enterochromaffin cells. Dim, dimension.





**Extended Data Figure 9 | Single-molecule FISH and immunofluorescence experiments confirm expression of markers for enteroendocrine cell subpopulations in the mouse small intestine.** **a, b**, Small intestine cryosections were hybridized with smFISH probe libraries. Scale bar, 10  $\mu\text{m}$ . **a**, The novel marker *Ucn3* was found in a small number of *Tac1*-positive cells. Probes against *Tac1*, conjugated with TMR (upper panel, red), and against *Ucn3*, conjugated with Cy5 (lower panel, green), were hybridized to the tissue sections. Dashed lines indicate cell borders. A cell co-expressing the two markers (*Tac1*<sup>+</sup>/*Ucn3*<sup>+</sup>) is shown in the left column. A cell expressing *Tac1*, but not *Ucn3* (*Tac1*<sup>+</sup>/*Ucn3*<sup>-</sup>), is shown in the right column. **b**, The novel marker *Vgf* is expressed by a subpopulation of *Tph1*-positive cells. Probes against *Tph1*, conjugated to TMR (upper panel, red), and against *Vgf*, conjugated to Cy5 (lower panel, green), were used for hybridization. Cell borders were stained by phalloidin-AlexaFluor 488 (not shown). Dashed lines demarcate cell borders. A *Tph1*-positive cell, expressing *Vgf* (*Tph1*<sup>+</sup>/*Vgf*<sup>+</sup>) is shown in the left column. An example of a *Tph1*-positive cell, expressing no *Vgf* (*Tph1*<sup>+</sup>/*Vgf*<sup>-</sup>) is shown in the right column. **c, d**, Immunostaining was

performed on cryosections of mouse small intestinal tissue. Scale bar, 20  $\mu\text{m}$ . **c**, Expression of UCN3 was observed in a few TAC1-positive cells within the jejunum. Frozen tissue sections were indirectly stained with anti-UCN3 (left panel, red), and anti-TAC1 (middle panel, light blue) antibodies. Nuclei were visualized with DAPI (dark blue). A cell, expressing both markers (TAC1<sup>+</sup>/UCN3<sup>+</sup>), is shown in the upper row. A cell, positive for TAC1, but not UCN3 (TAC1<sup>+</sup>/UCN3<sup>-</sup>), is shown in the lower row. The arrowhead points at the UCN3-negative cell. **d**, VGF is expressed by a subpopulation of TAC1-positive jejunal and ileal cells. VGF (left panel, red) and TAC1 (second panel from the left, grey) expression was visualized with indirect immunostaining. Expression of AFP was detected using a directly conjugated antibody (second panel from the right, green). Nuclei were counterstained with DAPI (blue). A TAC1-positive cell, expressing VGF and AFP (TAC1<sup>+</sup>/VGF<sup>+</sup>/AFP<sup>+</sup>) is shown in the upper panel. An example of a TAC1-positive cell, expressing no VGF or AFP (TAC1<sup>+</sup>/VGF<sup>-</sup>/AFP<sup>-</sup>) is shown in the lower panel. Arrowheads point at the VGF- and AFP-negative cell.



**Extended Data Figure 10 | Purification of *Lgr5*-GFP<sup>+</sup> cells from reporter mouse<sup>23</sup> organoids.** Single small intestinal organoid cells, derived from *Lgr5*-GFP mice were sorted by FACS. In total, 96 cells were sequenced from a single experiment on four lanes with 31,065,854 reads in total of which 33% could be mapped to the transcriptome. Every UMI-derived transcript was sequenced on average 6.4 times. A median number of 9,626 transcripts per cell were quantified. After filtering out cells with <3,000 transcripts and genes with <5 transcripts in all cells or >500 transcripts in a single cell, 74 cells remained for analysis. **a**, The t-SNE map shows the cell types identified by RaceID. Only a single predominant cell type and few outliers were observed. Cluster 1 comprises intestinal stem cells while the few outliers represent Paneth and enteroendocrine cells. **b–d**, The t-SNE maps show expression of the stem cell marker *Lgr5* (**b**), the stem cell marker *Lrig1* (**c**), and the +4 niche marker *Hopx*

(**d**). All markers are homogenously expressed across all cells at low transcript counts. We only observed marginal expression of the stem cell marker *Bmi1* in few cells and we did not observe expression of *Tert* in any of the cells, which is likely owing to the limited sensitivity of the method. The RaceID results indicate that *Lgr5*-positive intestinal stem cells represent a uniform population of cells. **e**, **f**, Combined analysis of random organoid and *Lgr5*-positive cells using RaceID. The initial clusters of the random organoid cells are conserved. The *Lgr5*-positive cells give rise to a uniform group that merges with the CBC/TA cluster from the random organoid cells (cluster 2). Shown is a heat map representation (**e**) and a t-SNE map (**f**) of the cell-to-cell transcriptome distances. Clusters are indicated by the same colours along the axes of the heat map (**e**) and for individual data points in the t-SNE map (**f**). Dim, dimension.



# Distinct EMT programs control normal mammary stem cells and tumour-initiating cells

Xin Ye<sup>1</sup>, Wai Leong Tam<sup>1,2,3</sup>, Tsukasa Shibue<sup>1</sup>, Yasemin Kaygusuz<sup>1</sup>, Ferenc Reinhardt<sup>1</sup>, Elinor Ng Eaton<sup>1</sup> & Robert A. Weinberg<sup>1,4,5</sup>

Tumour-initiating cells (TICs) are responsible for metastatic dissemination and clinical relapse in a variety of cancers<sup>1,2</sup>. Analogies between TICs and normal tissue stem cells have led to the proposal that activation of the normal stem-cell program within a tissue serves as the major mechanism for generating TICs<sup>3–7</sup>. Supporting this notion, we and others previously established that the Slug epithelial-to-mesenchymal transition-inducing transcription factor (EMT-TF), a member of the Snail family, serves as a master regulator of the gland-reconstituting activity of normal mammary stem cells, and that forced expression of Slug in collaboration with Sox9 in breast cancer cells can efficiently induce entrance into the TIC state<sup>8</sup>. However, these earlier studies focused on xenograft models with cultured cell lines and involved ectopic expression of EMT-TFs, often at non-physiological levels. Using genetically engineered knock-in reporter mouse lines, here we show that normal gland-reconstituting mammary stem cells<sup>9–11</sup> residing in the basal layer of the mammary epithelium and breast TICs originating in the luminal layer exploit the paralogous EMT-TFs Slug and Snail, respectively, which induce distinct EMT programs. Broadly, our findings suggest that the seemingly similar stem-cell programs operating in TICs and normal stem cells of the corresponding normal tissue are likely to differ significantly in their details.

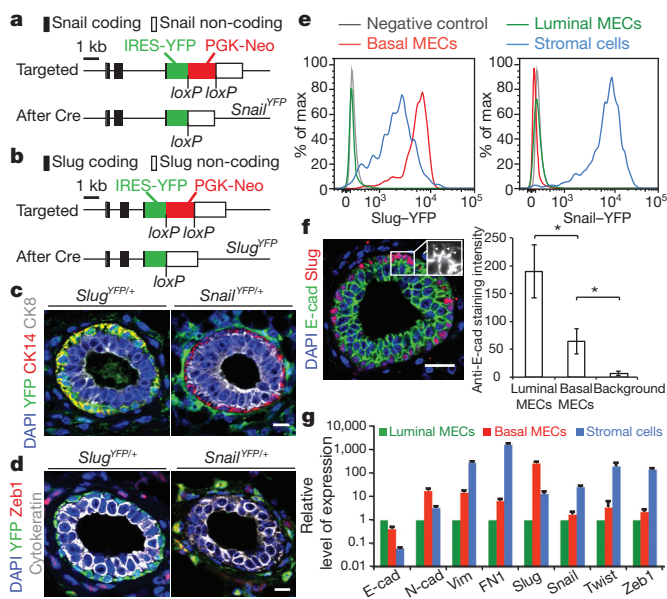
To define the functions of endogenously encoded, physiologically regulated Snail family EMT-TFs in breast cancer pathogenesis *in vivo*, we generated knock-in IRES (internal ribosomal entry site)–YFP (yellow fluorescent protein) reporters for *Slug* (also called *Snai2*) and *Snail* (also called *Snai1*) (Fig. 1a, b). These knock-in reporters faithfully reflected the expression of the endogenous genes (Extended Data Fig. 1a, b), and enabled the isolation of Slug<sup>+</sup> or Snail<sup>+</sup> cells by fluorescence-activated cell sorting (FACS) (Extended Data Fig. 6e–h).

Using these reporters, we found that Slug was expressed at higher levels in the normal mammary stem cell (MaSC)-enriched basal mammary epithelial cells (MECs) compared to the stromal fibroblasts surrounding the mammary ducts. In contrast, the EMT-TFs Snail, Twist and Zeb1 were expressed in stromal fibroblasts but not in either basal or luminal MECs (Fig. 1c–e, g and Extended Data Fig. 1c–f). In addition to the differential expression of EMT-TFs, the MaSC-enriched basal MECs displayed intermediate expression levels of both epithelial and mesenchymal markers (Fig. 1f, g and Extended Data Fig. 1g). Hence, Slug expression in the normal basal MECs was associated with only a partial conversion to the mesenchymal state.

Given the differential expression patterns of Slug and Snail, we undertook to analyse their expression during tumour development using the MMTV–PyMT transgenic model of mammary tumour formation, which mirrors the multi-step progression of human breast cancers beginning from hyperplastic lesions to high-grade carcinomas that spontaneously metastasize to the lungs<sup>12</sup>. In the initially formed hyperplastic lesions, we noted a marked reduction of the relative frequency of Slug–YFP<sup>+</sup> cells compared to normal glands, contrary to the hypothesis that activation of the Slug EMT-TF might be the preferred

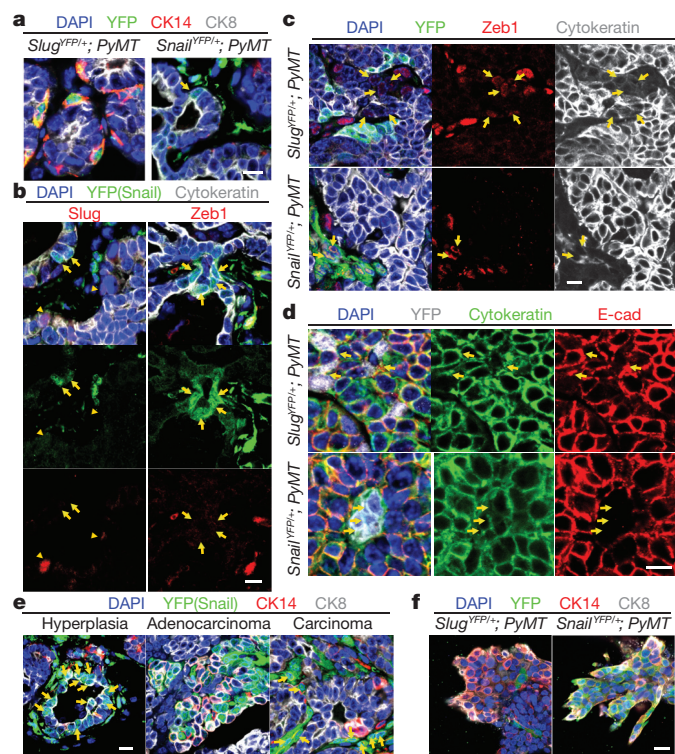
mechanism to generate TICs. These Slug–YFP<sup>+</sup> cells were cytokeratin14<sup>+</sup> (CK14) (Fig. 2a and Extended Data Fig. 2f), indicating that Slug expression was still confined to cells of the basal lineage, as was the case within the normal ducts. In these early-stage lesions, we detected for the first time Snail–YFP expression in a small fraction of the neoplastic cells displaying CK8<sup>+</sup>Slug<sup>−</sup>Zeb1<sup>−</sup> luminal characteristics (Fig. 2a, b and Extended Data Fig. 2a–c).

As these early-stage tumours progressed to high-grade carcinomas, the Slug<sup>+</sup> cells remained largely confined to the basal sectors of each epithelial island, whereas the Snail<sup>+</sup> cancer cells were sometimes fully detached from the epithelial islands and exhibited an elongated mesenchymal morphology (Fig. 2c). We found that virtually all Snail–YFP<sup>+</sup> tumour cells had lost E-cadherin expression and activated expression of the Zeb1 EMT-TF; in contrast, most Slug–YFP<sup>+</sup> tumour cells retained junctional E-cadherin and lacked Zeb1 expression



**Figure 1 | Differential expression of Slug and Snail in normal mammary glands.** **a, b**, Targeting strategies for the knock-in alleles. **c, d**, Normal mammary glands of the indicated genotypes were stained for the indicated proteins. **e**, FACS histograms showing relative expression levels of the YFP reporters in normal adult mammary cell subpopulations. **f**, Normal mammary gland stained for E-cadherin (E-cad) and Slug. Arrowheads indicate the junctions between basal MECs. Right panel: quantifications of Anti-E-cadherin staining intensities at the junctions between luminal MECs and basal MECs in a representative mammary gland (mean  $\pm$  s.d.,  $n = 20$ , cell junctions,  $*P < 0.00001$ ). Data represent analyses of six glands. **g**, Representative qRT-PCR quantification of the indicated EMT markers (mean  $\pm$  s.e.m., technical triplicates). Levels in luminal MECs were set to one. Data represent three independent experiments. All scale bars indicate 20  $\mu$ m.

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore. <sup>3</sup>Cancer Science Institute of Singapore, 14 Medical Drive, Singapore 117599, Singapore. <sup>4</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Ludwig Center for Molecular Oncology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.



**Figure 2 | Differential expression of Slug and Snail in mammary tumours.** **a, b,** Hyperplastic mammary lesions of the indicated genotypes were stained for the indicated proteins. Arrow in **a** indicates Snail–YFP and CK8 double-positive cells. Arrows and arrowheads in **b** indicate Snail–YFP and cytokeratin double-positive cells, and Slug-positive cells, respectively. **c, d,** High-grade carcinomas of the indicated genotypes were stained for the indicated proteins. Arrows indicate Zeb1 and cytokeratin double-positive cells (**c**) and the junctions between YFP-positive carcinoma cells (**d**). **e,** *Snail*<sup>YFP/+</sup>; MMTV-PyMT tumours were stained for the indicated proteins. Arrows indicate Snail–YFP-positive carcinoma cells. **f,** Tumour organoids of the indicated genotypes were stained for the indicated proteins. Images represent three independent experiments. All scale bars indicate 10  $\mu$ m.

(Fig. 2c, d and Extended Data Fig. 2d). Therefore, Snail rather than Slug is associated with more complete expression of mesenchymal traits in mammary tumours.

Notably, as tumours progressed, we noted that the *Snail*–YFP<sup>+</sup> cells gradually acquired basal CK14 expression and lost luminal CK8 expression (Fig. 2e and Extended Data Fig. 2c, e), consistent with the proposal that in human breast carcinomas, aggressive cancer cells exhibiting basal features can arise from luminal precursors<sup>13–17</sup>. To compare the activation of Slug and Snail during such luminal–basal transitions, we used an organoid culture system in which CK14 is spontaneously activated as the tumour cells invade into a type I collagen gel<sup>13</sup>. We dissociated adenocarcinomas into tumour organoids as previously described<sup>13</sup>. These freshly isolated tumour organoids were almost exclusively of luminal phenotype (CK8<sup>+</sup>CK14<sup>−</sup>) and lacked both Slug–YFP and Snail–YFP expression (Extended Data Fig. 3a). After 48 h in culture, CK14 expression was induced in tumour cells at the invasive fronts of these organoids. Strikingly, this induction was tightly associated with Snail–YFP but not Slug–YFP activation (Fig. 2f and Extended Data Fig. 3b, c).

Taken together, these analyses indicate that the EMT-TF that is activated in the MMTV-PyMT mammary tumours (that is, Snail) is distinct from the one expressed in the normal gland-reconstituting MaSCs (that is, Slug), and that expression of Snail rather than its paralogue Slug is associated with potent EMT activation and eventual acquisition of basal features (Extended Data Fig. 3d).

To extend and generalize these observations further, we compared Slug, Snail and Zeb1 expression patterns in the MMTV-Neu<sup>18</sup> and

BRCA1/p53-minus<sup>19</sup> transgenic models of breast cancer development. Snail activation had previously been associated with recurrence and metastasis formation in the MMTV-Neu model<sup>20,21</sup>. Consistently, we found that Snail and Zeb1 are activated in ~1–2% of tumour cells within MMTV-Neu tumours that are associated with metastatic disease, which contained no evidence of Slug expression (Extended Data Fig. 4a–c). The BRCA1/p53-minus tumours were highly heterogeneous and harboured areas with epithelial and glandular phenotypes as well as areas with mesenchymal and invasive phenotypes (Extended Data Fig. 4d, g). We found that Slug was predominantly expressed by cells occupying the basal sectors of the gland-like structures (echoing its expression in normal ducts), whereas Snail and Zeb1 were predominantly expressed in the invasive cells (Extended Data Fig. 4d–h). Hence, differential expression of Slug and Snail appears to be a conserved feature of mammary tumours driven by diverse oncogenic signals.

We were curious whether differential expression of SLUG and SNAIL was also observable in human breast cancer cell lines. Mirroring our observations in the mouse models, SNAIL but not SLUG was expressed in luminal breast cancer cells, and the two EMT-TFs were expressed in distinct (but occasionally overlapping) populations in basal breast cancer cells (Extended Data Fig. 5a, b). In addition to these genetically unrelated breast cancer cell lines, we compared SLUG and SNAIL expression in an MCF10A-based model of breast cancer progression<sup>22</sup>. This model consists of the immortalized, non-tumorigenic MCF10A human MECs, MCF10A cells transformed with an *H-RAS* oncogene (MCF10A-Ras), and a cell line established from an MCF10A-Ras cell-derived carcinoma (MCF10A-Ras-C). We found that SLUG was expressed in MCF10A cells, but was downregulated in MCF10A-Ras-C cells. In contrast, SNAIL was absent in MCF10A cells, but underwent activation in MCF10A-Ras cells and further upregulation in MCF10A-Ras-C cells (Extended Data Fig. 5c, d).

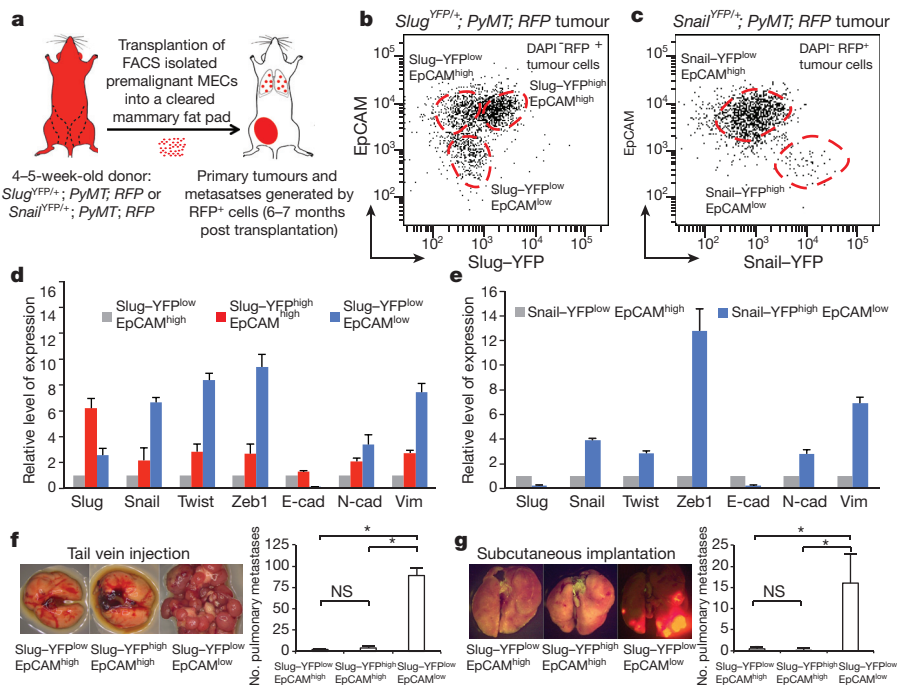
To summarize, these results indicated that the differential expression of Slug and Snail is a common feature of mammary tumours despite their different subtypes, genetic backgrounds and oncogenic drivers. Although Snail is absent in normal MECs, it often becomes activated during breast cancer progression. Indeed, SNAIL expression was detected in ~80% microdissected human invasive ductal carcinomas<sup>23</sup>.

Our observations raised the question of whether the *Slug*<sup>+</sup>, *Snail*<sup>+</sup>, or yet other cell subpopulations within mammary tumours were enriched in TICs. To address this issue, we developed a system that allowed us to isolate these various subpopulations. Thus, we FACS-purified premalignant EpCAM<sup>+</sup> MECs from 4–5-week-old *Slug*<sup>YFP/+</sup>; MMTV-PyMT;RFP (red fluorescent protein) or *Snail*<sup>YFP/+</sup>; MMTV-PyMT;RFP animals, and thereafter implanted these cells into cleared mammary fat pads of hosts that lacked these transgenes (Fig. 3a). The implanted premalignant MECs first grew as rudimentary ductal structures and then progressed over 6–7 months to form high-grade carcinomas that metastasized to the lungs (Extended Data Fig. 6a, b).

We FACS-resolved the RFP<sup>+</sup> carcinoma cells based on the expression levels of the YFP (Slug or Snail) reporter and the EpCAM epithelial marker. In the high-grade carcinomas and corresponding pulmonary metastases, EpCAM expression was downregulated in 4–12% of the carcinoma cells. These EpCAM<sup>low</sup> cells had low Slug–YFP expression (Fig. 3b and Extended Data Fig. 5c) but high Snail–YFP expression (Fig. 3c and Extended Data Fig. 5d). In contrast, the EpCAM<sup>high</sup> populations were *Snail*–YFP<sup>low</sup> (Fig. 3c), and were composed of *Slug*–YFP<sup>low</sup>EpCAM<sup>high</sup> and *Slug*–YFP<sup>high</sup>EpCAM<sup>high</sup> subpopulations (Fig. 3b). Hence, EpCAM expression was inversely correlated only with Snail expression.

Using quantitative reverse transcription PCR (qRT-PCR) analyses, we confirmed that *Snail* expression was highest in the EpCAM<sup>low</sup> subpopulations, whereas *Slug* was enriched in EpCAM<sup>high</sup> subpopulations (Fig. 3d, e). As expected, strong induction of mesenchymal markers and suppression of E-cadherin were only seen in the





**Figure 3 | Breast TICs express Snail.** **a**, Schematic of the transplantation model. **b**, **c**, FACS profiles of the high-grade carcinomas of the indicated genotypes. **d**, **e**, Representative qRT-PCR analyses of the expressions of EMT markers (mean + s.e.m., technical triplicates) in indicated subpopulations of high-grade carcinomas derived from *Slug*<sup>YFP/+</sup>; *MMTV*-*PyMT*; *RFP* (d) and *Snail*<sup>YFP/+</sup>; *MMTV*-*PyMT*; *RFP* MECs (e). Expression levels in *Slug*-YFP<sup>low</sup> EpCAM<sup>high</sup> cells and *Snail*-YFP<sup>low</sup> EpCAM<sup>high</sup> cells were set to one. Data in **d** and **e** represent three independent experiments. **f**, **g**, Metastatic outgrowths generated by indicated subpopulations after tail vein injection. \**P* < 0.0001 (**f**) and after subcutaneous implantation \**P* = 0.019 (**g**) (mean + s.d., *n* = 5 animals per group). NS, not significant. Data in **f** and **g** represent four and three independent experiments, respectively. There is Source Data associated with this figure.

EpCAM<sup>low</sup> subpopulations (Fig. 3d, e). Using cell lines derived from *Slug*<sup>YFP/+</sup>; *MMTV*-*PyMT* and *Snail*<sup>YFP/+</sup>; *MMTV*-*PyMT* tumours, we also observed the segregation of *Slug* and *Snail* expression at the protein level and associated *Snail* but not *Slug* expression with strong induction of a mesenchymal phenotype (Extended Data Fig. 6e–h).

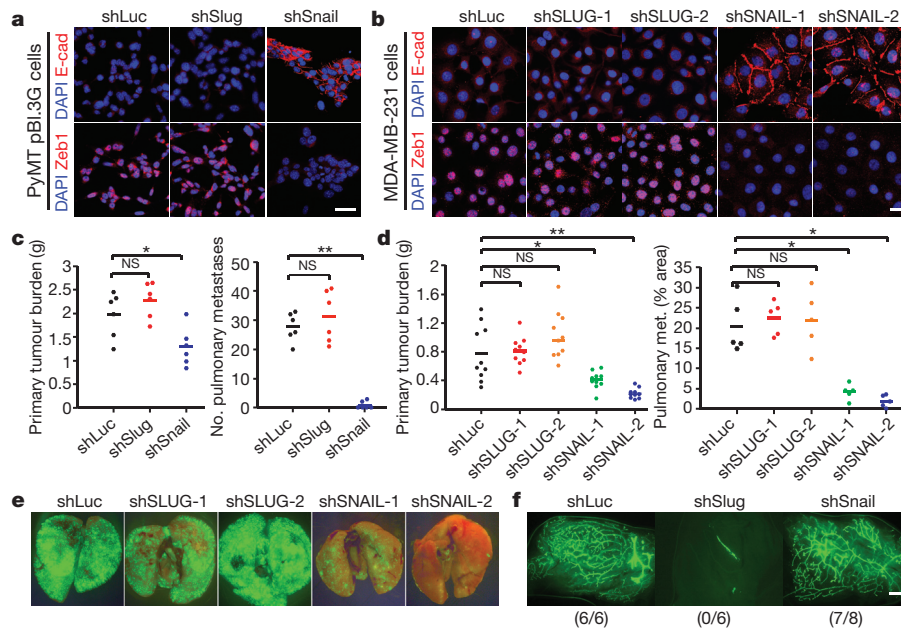
Together, these observations demonstrated our ability to resolve and isolate distinct tumour cell subpopulations with high levels of either *Slug* (*Slug*<sup>high</sup>) or *Snail* (*Snail*<sup>high</sup>) from the same primary tumours (Extended Data Fig. 7a), allowing us, in turn, to directly compare their respective tumour-initiating activities. To this end, we FACS-purified these subpopulations (Extended Data Fig. 6e–h) and implanted each at limiting dilutions to score tumour formation. Overall, *Snail*<sup>high</sup> subpopulations exhibited more than two orders of magnitude higher proportions of TICs than did the other subpopulations. In contrast, the *Slug*-YFP<sup>high</sup> cells were as deficient in tumour-initiating ability as the *Slug*-YFP<sup>low</sup> EpCAM<sup>high</sup> cells (Extended Data Fig. 7b, c). Hence, *Snail* but not *Slug* was tightly associated with a TIC phenotype.

To compare the TIC activities and metastatic powers of these tumour cell subpopulations coexisting in the same primary tumours *in vivo*, we FACS-purified each subpopulation from highly metastatic carcinomas generated by *Slug*<sup>YFP/+</sup>; *MMTV*-*PyMT*; *RFP* cells and introduced them via the tail vein to gauge their respective abilities to seed pulmonary metastases. Notably, the *Snail*<sup>high</sup> cells consistently gave rise to far more metastatic outgrowths relative to the EpCAM<sup>high</sup> subpopulations. In particular, 40,000 *Snail*<sup>high</sup> cells from a highly metastatic primary tumour (Extended Data Fig. 8a) seeded on average ~90 large metastases in each animal. In contrast, 40,000 cells of the *Slug*<sup>high</sup> and *Slug*-YFP<sup>low</sup> EpCAM<sup>high</sup> subpopulations from the same tumour seeded an average of only 3.6 and 2.2 metastases per animal, respectively (Fig. 3f and Extended Data Fig. 8b). The *Snail*<sup>high</sup> cells were also far more metastatic than the other two populations when implanted subcutaneously (Fig. 3g and Extended Data Fig. 8e–g). Interestingly, the metastatic outgrowths formed by the *Snail*<sup>high</sup> cells harboured gland-like structures composed of both *Slug*<sup>+</sup> and *Slug*<sup>−</sup> cells (Extended Data Fig. 8c, d). Hence, the *Snail*<sup>high</sup> cells that seeded metastases were capable of differentiating within these outgrowths, thereby regenerating the complex cellular hierarchy present in the original primary tumours. These data also revealed that the TICs did not derive from basal MaSC-like cells (that is, the *Slug*<sup>high</sup> cells) but instead arose in a different cell population.

We were curious whether correlates of these distinct behaviours of *SLUG* and *SNAIL* could be found in human clinical data sets, and therefore examined the prognostic powers of *SLUG* or *SNAIL* (ref. 24). Across various patient populations, we found that only elevated expression of *SNAIL* consistently associated with poor survival (Extended Data Fig. 9a).

These results strongly argue for divergent roles of *Slug* and *Snail*, and predicted that shutdown of *Snail* could selectively eliminate breast TICs. To test this notion, we knocked down either *Slug* or *Snail* in advanced *MMTV*-*PyMT* carcinoma-derived pBI.3G cells and MDA-MB-231 human breast cancer cells (Extended Data Fig. 9b, c). Strikingly, *Snail* knockdown but not *Slug* knockdown induced mesenchymal-to-epithelial transition (MET), leading to loss of *Zeb1* and reactivation of *E-cadherin* (Fig. 4a, b). When these cells were injected orthotopically into mammary fat pads, *Snail* but not *Slug* knockdown attenuated primary tumour growth and strongly suppressed their metastatic spreading (Fig. 4c–e). Similarly, across a panel of human breast cancer cell lines, we found that *SNAIL* knockdown significantly suppressed tumour initiation in most of them, while *SLUG* knockdown failed to do so (Extended Data Fig. 9d–f). In contrast to the responses of breast cancer cells, the organoid-forming and gland-reconstituting activities of normal murine MaSCs were markedly affected by *Slug* knockdown but not by *Snail* knockdown (Fig. 4f and Extended Data Fig. 9g).

Given the distinct functions exerted by *Slug* and *Snail*, we analysed the transcription programs controlled by these paralogous EMT-TFs using ChIP-seq. We focused on two *MMTV*-*PyMT* tumour cell lines that differed in *Slug* and *Snail* expression and tumorigenic potential (Extended Data Fig. 10a, b). We recovered similar numbers of chromatin regions that were enriched for either *Slug* or *Snail* binding (Fig. 5a). Across the genome, *Slug*- and *Snail*-binding sites displayed similar fold enrichments and were both enriched for the known *Snail* family recognition CANNTG E-box motif<sup>25</sup> (Extended Data Fig. 10c, d). We found that *Snail* occupied 10,129 promoters, far more than that occupied by *Slug* (2,475 promoters) (Fig. 5b). Interestingly, the promoters of genes encoding key mesenchymal markers were only bound by *Snail* but not by *Slug* (Extended Data Fig. 10e). Gene-set enrichment analyses (GSEA) confirmed that *Snail*-bound but not *Slug*-bound genes were significantly enriched for EMT-related signatures (Fig. 5c). In particular, *Snail*, but not *Slug*, occupied the promoter of *Zeb1* (Fig. 5d), a master regulator of



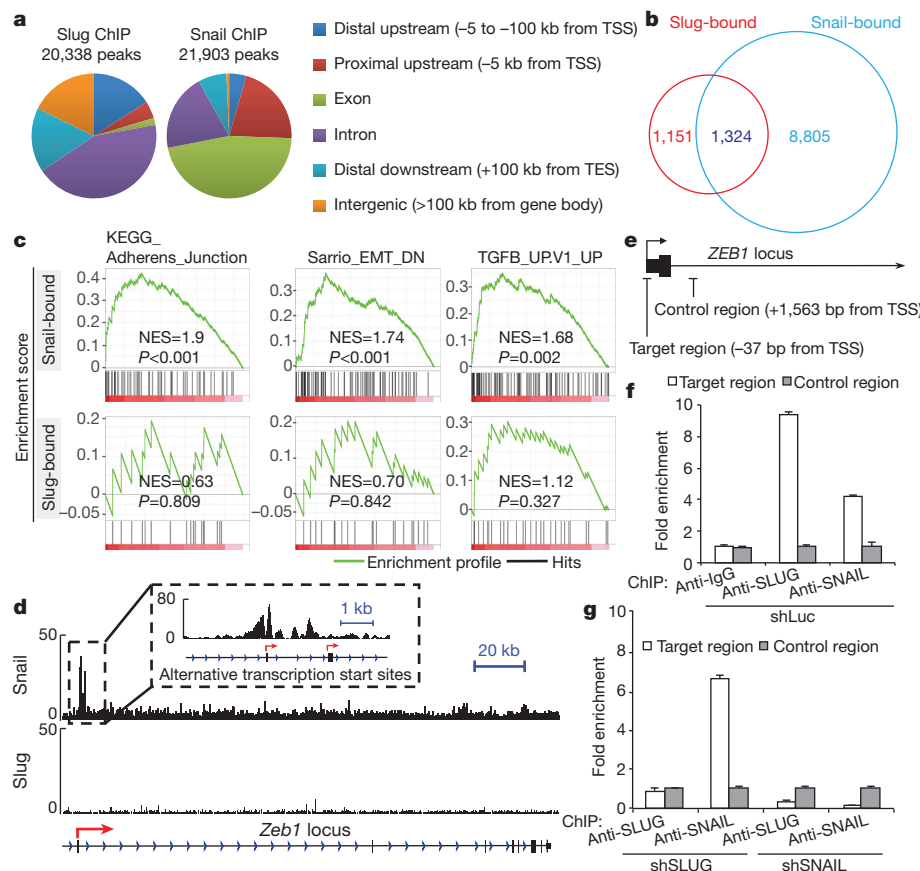
**Figure 4 | Depletion of Snail selectively affects breast TICs.** **a, b**, Immunofluorescence images of the shRNA-transduced pBL3G cells (**a**) and MDA-MB-231 cells (**b**). Scale bars, 20  $\mu$ m. **c, d**, Primary tumour burdens and pulmonary metastases formed by orthotopically implanted pBL3G cells (**c**, unilateral implantation,  $n = 5$  animals per group,  $*P = 0.026$ ,  $**P = 1.9 \times 10^{-5}$ ) and MDA-MB-231 cells (**d**, bilateral implantation,  $n = 5$  animals per group,  $*P < 0.01$ ,  $**P < 0.001$ ). NS, not significant. Source Data is associated with this figure. **e**, Fluorescent images of whole-mount lungs showing spontaneous metastases formed by the orthotopically implanted GFP-labelled MDA-MB-231 cells. **f**, Whole-mount fluorescent images of the mammary fat pads implanted with the indicated GFP-expressing primary murine MECs. Scale bar, 1 mm.

TICs under a variety of settings<sup>26–28</sup>. The binding of Snail appears to activate *Zeb1* expression in mammary tumour cells, because knockdown of Snail led to downregulation of *Zeb1* (Fig. 4a, c), and ectopic expression of SNAIL in human MECs induced *ZEB1* expression (Extended Data Fig. 10f). Indeed, Snail had been reported to activate *Zeb1* expression in non-mammary types of epithelial cells<sup>29,30</sup>.

To investigate the possible differential abilities of SLUG and SNAIL in controlling *ZEB1* expression in human breast cancer cells, we used ChIP-qPCR to examine SLUG and SNAIL binding at the *ZEB1* promoter in MDA-MB-231 cells, which co-express SLUG and SNAIL.

Notably, although both SLUG and SNAIL appeared to occupy the *ZEB1* promoter in these cells, SLUG binding (but not SLUG expression) was diminished in SNAIL knockdown cells (Fig. 5e–g). Since the knockdown of SNAIL, but not SLUG, resulted in downregulation of *ZEB1* in MDA-MB-231 cells (Fig. 4c) and the binding of SLUG to the *ZEB1* promoter is dependent on SNAIL expression, we concluded that *ZEB1* expression was controlled by SNAIL but not SLUG in MDA-MB-231 cells as well.

Our data underscore profound differences in the transcription-regulating activities of the endogenously encoded Slug and Snail



**Figure 5 | Slug and Snail control different targets.** **a**, Distribution of summits of all Snail ChIP-seq and Slug ChIP-seq peaks. TES, transcription end site; TSS, transcription start site. **b**, Venn diagram showing the numbers of promoters occupied by Slug and Snail. **c**, GSEA analyses of published EMT-related data sets for Snail-bound and Slug-bound genes. NES, normalized enrichment score. **d**, ChIP-seq signals for Slug and Snail at the *Zeb1* locus. Arrows indicate TSSs. **e–g**, Fold-enrichment of SLUG and SNAIL binding at the *ZEB1* promoter relative to background measured by ChIP-qPCR (mean  $\pm$  s.e.m., technical triplicates). Data represent two independent experiments.



EMT-TFs, and provide indications that normal stem cells and TICs of the same tissue-of-origin could arise from different cellular compartments and exploit different molecular signalling circuits to activate related but distinct signalling pathways. We have previously correlated high levels of SLUG expression with poor prognosis in human breast cancer patients, and attributed such correlation to the experimentally observed EMT-inducing function of SLUG<sup>8</sup>. However, on the basis of the present work, we propose that the prognostic power of SLUG expression may be due in large part to its strong association with basal differentiation, which is, on its own, a well-known feature of aggressive breast cancers<sup>16</sup>. Nonetheless, since our *in vivo* analyses focused on MMTV promoter-driven tumours, which appear to derive primarily from luminal MECs, we suggest that the functions of Slug in basal MEC-derived tumours remain to be further characterized.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 December 2014; accepted 9 July 2015.**

**Published online 2 September 2015.**

- Al-Hajj, M. & Clarke, M. F. Self-renewal and solid tumor stem cells. *Oncogene* **23**, 7274–7282 (2004).
- O'Brien, C. A., Kreso, A. & Dick, J. E. Cancer stem cells in solid tumors: an overview. *Semin. Radiat. Oncol.* **19**, 71–77 (2009).
- Visvader, J. E. & Lindeman, G. J. Cancer stem cells: current status and evolving complexities. *Cell Stem Cell* **10**, 717–728 (2012).
- Chaffer, C. L. *et al.* Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc. Natl Acad. Sci. USA* **108**, 7950–7955 (2011).
- Gupta, P. B. *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
- Beck, B. & Blanpain, C. Unravelling cancer stem cell potential. *Nature Rev. Cancer* **13**, 727–738 (2013).
- Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).
- Guo, W. *et al.* Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* **148**, 1015–1028 (2012).
- Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
- Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
- Prater, M. D. *et al.* Mammary stem cells have myoepithelial cell properties. *Nature Cell Biol.* **16**, 942–950 (2014).
- Lin, E. Y. *et al.* Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am. J. Pathol.* **163**, 2113–2126 (2003).
- Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155**, 1639–1651 (2013).
- Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nature Med.* **15**, 907–913 (2009).
- Molyneux, G. *et al.* BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* **7**, 403–417 (2010).
- Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* **23**, 2563–2577 (2009).
- Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
- Muller, W. J., Sinn, E., Pattengale, P. K., Wallace, R. & Leder, P. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. *Cell* **54**, 105–115 (1988).
- Xu, X. *et al.* Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nature Genet.* **22**, 37–43 (1999).
- Moody, S. E. *et al.* The transcriptional repressor Snail promotes mammary tumor recurrence. *Cancer Cell* **8**, 197–209 (2005).
- Tran, H. D. *et al.* Transient SNAIL1 expression is necessary for metastatic competence in breast cancer. *Cancer Res.* **74**, 6330–6340 (2014).
- Papageorgis, P. *et al.* Smad signaling is required to maintain epigenetic silencing during breast cancer progression. *Cancer Res.* **70**, 968–978 (2010).
- Natrajan, R. *et al.* An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. *Breast Cancer Res. Treat.* **121**, 575–589 (2010).
- Gyorffy, B. *et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**, 725–731 (2010).
- Nieto, M. A. The snail superfamily of zinc-finger transcription factors. *Nature Rev. Mol. Cell Biol.* **3**, 155–166 (2002).
- Chaffer, C. L. *et al.* Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* **154**, 61–74 (2013).
- Siebzehnubel, F. A. *et al.* The ZEB1 pathway links glioblastoma initiation, invasion and chemoresistance. *EMBO Mol. Med.* **5**, 1196–1212 (2013).
- Wellner, U. *et al.* The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nature Cell Biol.* **11**, 1487–1495 (2009).
- Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
- Guaita, S. *et al.* Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. *J. Biol. Chem.* **277**, 39209–39216 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The pBl.1 and pBl.3 murine PyMT tumour cell lines, from which pBl.1G and pBl.3G were derived, were gifts from the Harold L. Moses laboratory. We thank G. Bell for helping analysing the ChIP-seq data. We thank R. Bronson for help assessing the histopathology of the murine tumours. We thank A. Lambert and S. Thiagalingam for providing the MCF10A-Ras and MCF10A-Ras-C cells. We thank W. Israelsen and M. Vander Heiden for providing the MMTV-cre;p53<sup>+/−</sup>;BRCA1<sup>fl/fl</sup> murine mammary tumour samples. We thank the Keck Microscopy Facility at the Whitehead Institute for microscopy assistance and the Koch Institute Swanson Biotechnology Center (SBC) for technical support, especially the Histology and ES cell and Transgenics Cores. R.A.W. is an American Cancer Society and Ludwig Foundation professor. W.L.T. is supported by the National Research Foundation, Singapore (NRF-NRFF2015-04). This research was supported by the Breast Cancer Research Foundation, the Samuel Waxman Cancer Research Foundation, the Ludwig Center for Molecular Oncology at MIT, National Cancer Institute Program P01-CA080111, R01-CA078461, U01-CA184897 (to R.A.W.), K99-CA194160 (to X.Y.), the Wilshire Charitable Foundation/Andria and Paul Heafy Postdoctoral Fellowship (to X.Y.), the Mattina R. Proctor Foundation, and the Helen Hay Whitney Foundation (to X.Y.).

**Author Contributions** X.Y. and R.A.W. conceived the project and prepared the manuscript. X.Y. designed and performed the experiments and analysed the data. W.L.T. contributed to the ChIP-seq experiments. T.S. generated the knock-in animals. Y.K. performed experiments and quantifications. F.R. performed the mammary fat pad injections. E.E. provided technical support.

**Author Information** The ChIP-seq results have been deposited to Gene Expression Omnibus (GEO) under accession number GSE61198. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.Y. ([xye@wi.mit.edu](mailto:xye@wi.mit.edu)) or R.A.W. ([weinberg@wi.mit.edu](mailto:weinberg@wi.mit.edu)).

## METHODS

**Animals.** The *Slug*<sup>YFP</sup> and *Snail*<sup>YFP</sup> alleles were generated by homologous recombination in mouse embryonic stem (ES) cells using standard gene-targeting methods. The MMTV-Neu animals and CAG-mRFP animals were obtained from the Jackson Laboratory (stock numbers 005038 and 005884). The MMTV-PyMT animals were originally obtained from the Jackson Laboratory (stock number 002374) and backcrossed for five generations to C57BL/6 background. Mice were housed and handled in accordance with protocols approved by the Animal Care and Use Committees of the Massachusetts Institute of Technology. Animals were randomized by age and weight. The investigators were blinded to allocation during experiments and outcome assessment for experiments shown in Fig. 3f, g and Extended Data Fig. 10b.

**Genotyping.** PCR primers (5' to 3') for *Slug*<sup>YFP</sup> genotyping were (sense strand, AACCTTCTCCAGAATGTCGCTTCTG; antisense strand, TGCAGGTGTATC TTATACACGTGGC) and for *Snail*<sup>YFP</sup> genotyping were (sense strand, CTCCCG CATGTCCTTGCTCCACAAG; antisense strand, same as for *Slug*<sup>YFP</sup>). DNA extractions and subsequent PCR reactions were performed using the REDExtract-N-Amp Tissue PCR kit (Sigma XNAT). PCR was performed with 35 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, and elongation at 72 °C for 30 s.

**Tumour categorization.** MMTV-PyMT tumours were categorized into hyperplastic lesions, adenocarcinomas and high-grade carcinomas following the histological criteria described by Lin *et al.*<sup>12</sup>. Briefly, in the genetic background of our animals, hyperplastic lesions usually occur at about 8–10 weeks of age, and consist of regionally packed lobules formed on the duct. We collectively refer to the adenoma/MIN and early carcinoma stages defined by Lin *et al.*<sup>12</sup> as adenocarcinomas, as the tumours from 2.5 to 4-month-old animals often have mixed characteristics of both kinds. High-grade carcinoma corresponds to the late carcinoma/advanced invasive carcinoma as defined by Lin *et al.*<sup>12</sup>, and is always associated with metastatic disease.

**Immunofluorescence staining.** Tumours were fixed in 10% neutral buffered formalin overnight and embedded in paraffin for sectioning. Sections were cut at 5 µm. Tumour sections were deparaffinized in HistoClear II, and antigen retrieval was performed with Nuclear Decloaker (Biocare Medical) using a microwave. Sections were then blocked with 0.5% normal donkey serum (Jackson ImmunoResearch Laboratories) in PBST (PBS + 0.3% Triton X-100) for 1 h at room temperature. Sections were incubated with primary antibody at 4 °C overnight. After three washes with PBS, sections were incubated with secondary antibodies (Biotium) and DAPI for 2 h at room temperature, washed three times with PBS, and mounted in Prolong gold antifade reagent (Invitrogen P36930). For anti-Slug, anti-Snail and anti-Zeb1 immunofluorescence, the signals were amplified with the TSA Plus Systems (Perkin Elmer) following manufacturer's instructions.

Tumour organoids were fixed in collagen I gel with 4% paraformaldehyde for 1 h at room temperature, blocked with 0.5% normal donkey serum in PBST, and then incubated with primary antibodies at 4 °C overnight. After five washes with PBST, organoids were incubated with DAPI, secondary antibodies and Phalloidin at 4 °C overnight. After five washes with PBST, the collagen I gel containing the organoids were mounted in Prolong gold antifade reagent.

Cultured tumour cells were fixed in 4% paraformaldehyde and blocked with 0.1% normal donkey serum in PBST for half an hour at room temperature. Cells were incubated in specific primary antibodies for 1–2 h, washed three times with PBS, then incubated with secondary antibodies for 1 h at room temperature. After three washes with PBS, stained cells were mounted in Prolong gold antifade reagent.

Immunostained samples were imaged using Zeiss LSM710 and Zeiss LSM700 confocal microscopes and analysed with Zen software.

Antibodies used in this study are listed in the Supplementary Information.

**Tumour dissociation, FACS fractionation and derivation of tumour cell lines.** Tumours were taken from the animals aseptically. At least one fragment from each tumour was saved for histological staging of the tumour. The remainder of each tumour was then minced with a razor blade, and the minced chunks were then rinsed three times with PBS, and digested with collagenase A followed by trypLE-select (Invitrogen). The dissociated tumour cells were then washed twice with DMEM with 10% FBS, and filtered through a 70 µm and 40 µm cell strainer. The resulting cells were stained with DAPI, anti-EpCAM antibody, washed three times with PBS, and resuspended in PBS for flow cytometry analysis and FACS fractionation.

To establish tumour cell lines,  $1 \times 10^7$  dissociated and filtered tumour cells were plated in a 10 cm dish in DMEM/F12 supplemented with 5% adult bovine serum, non-essential amino acids (Invitrogen), and penicillin/streptomycin. On the next day, dead cells were removed by medium change, and the attached cells were passaged at 1:2 to 1:3 for about five passages until each culture was established.

Tumour cell lines were all derived from high-grade carcinomas of 6–7 months old females.

**Cell lines and cell culture.** PyMT tumour cell lines were cultured in DMEM/F12 (1:1) supplemented with 5% adult bovine serum (Sigma B9433), non-essential amino acid (Invitrogen 11140), and Pen/Strep (Invitrogen 15070). MDA-MB-231, MDA-MB-361 and MDA-MB-468 cells were culture in DMEM/F12 (1:1) supplemented with 10% inactivated fetal bovine serum (Sigma F4135) and Pen/Strep. SKBR3, BT20 and MCF7/Ras cells were cultured in DMEM supplemented with 10% inactivated fetal bovine serum and Pen/Strep. SUM149 and SUM159 cells were cultured in F12 media supplemented with 5% inactivated fetal bovine serum, insulin ( $5 \mu\text{g ml}^{-1}$ ), and hydrocortisone ( $1 \mu\text{g ml}^{-1}$ ). T47D, HS578T and BT549 cells were cultured in RPMI supplemented with 10% inactivated fetal bovine serum and Pen/Strep. MCF10A cells were cultured in DMEM/F12 supplemented with 5% horse serum (Sigma H1207), insulin ( $10 \mu\text{g ml}^{-1}$  Sigma I9278), EGF ( $100 \mu\text{g ml}^{-1}$  Sigma E9644), hydrocortisone ( $0.5 \text{ mg ml}^{-1}$  Sigma H0888), and cholera toxin ( $100 \text{ ng ml}^{-1}$  Sigma C8052). All human breast cancer cell lines were obtained from ATCC, and are free of mycoplasma contamination.

MMTV-PyMT tumour organoids culture was performed as previously described<sup>13</sup>. 3D organoid culture of primary MECs was performed as previously described<sup>8</sup>. Tumour-sphere assays were performed as described in Tam *et al.*<sup>31</sup>. 200 cells (1,000 cells for MDA-MB-361 cells) were plated into each well of 96-well ultra-low attachment plate to score tumour-sphere formation efficiency. Five wells were analysed for each condition.

**Tumour cell implantation.** For cleared mammary fat pad transplantation, primary MECs were isolated from the mammary glands of 4- to 5-week-old virgin female mice by collagenase and trypsin digestion. Sorted EpCAM<sup>+</sup> cells were suspended in 10 µl PBS containing 50% Matrigel, and injected into the inguinal mammary fat pads of 3-week-old NOD/SCID mice, whose endogenous mammary epithelium were cleared at the time of the injection. For tail-vein injection, tumour cells were resuspended in 100 µl PBS, and injected into male animals. The lungs were examined 4–8 weeks post injection. For subcutaneous injections, tumour cells were suspended in 100 µl PBS with or without Matrigel (as indicated in the figures) into the flanks of male animals. The tumour incidence and weight were measured 2–3 months post injection. For orthotopic tumour transplantations, tumour cells were resuspended in 20 µl media with the indicated amount of Matrigel and injected into mammary fat pads. Host animals were randomized by weight. In Fig. 4c,  $10^5$  PyMT tumour cells were injected unilaterally into the mammary fat pad (without matrigel) to score primary tumour burden and metastatic dissemination (six animals for each group). In Fig. 4d,  $10^6$  MDA-MB-231 cells were injected bilaterally (with 20% matrigel) to score primary tumour burden and metastatic dissemination (five animals for each group). For the metastasis quantification, the lungs were examined under a Leica fluorescence dissecting microscope. Blinded quantifications were performed when scoring the numbers of lung metastases. For limiting dilution analyses, the frequency of TICs in the cell population being transplanted was calculated using the Extreme Limiting Dilution Analysis Program (<http://bioinf.wehi.edu.au/software/elda/index.html>)<sup>32</sup>.

**Cleared mammary fat pad injection.** Cleared mammary fat pad injections were performed as previously described<sup>8</sup>. Briefly,  $1 \times 10^5$  cells were suspended in 10 µl PBS containing 50% Matrigel and injected into the inguinal mammary fat pads of NOD/SCID female mice that had previously been cleared of endogenous mammary epithelium. Gland reconstitutions were assessed under fluorescence dissecting microscope at 3 months post injection.

**RNA isolation, reverse transcription and qPCR analysis.** Total RNA from freshly sorted primary tumour cells was extracted using Trizol (Invitrogen), and column-purified with PicoPure RNA Isolation Kit (Applied Biosystems). cDNA synthesis was performed with 0.2–2 µg of total RNA using SuperScript III First-Strand Synthesis System (Invitrogen). mRNA levels were measured with gene-specific primers using the Roche LightCycler 480 system (Roche). Relative expression levels were normalized to  $\beta$ -actin. Primers used for qPCR analysis are listed in Supplementary Information.

**shRNA vectors.** The sources or targeting sequences of the shRNA used in the study are as follows: mouse shSlug, shSlug4 from Guo *et al.*<sup>8</sup>; mouse shSnail, from Shibue *et al.*<sup>33</sup>; human shSLUG, clones 1 and 2 from ref. 34; human shSNAIL, no. 1 TTCCTTGTTGCAGTATTG, no. 2 ATAAATACCAAGTGTA CCTT; shLuciferase, CCTAAGGTTAAGTCGCCCTCG.

**Meta-analysis of oncogenomic data.** To test whether the expression of *Slug* or *Snail* correlated with distant metastasis-free survival (all subtypes,  $n = 1,610$ ) and relapse-free survival (ER<sup>+</sup> patients,  $n = 1,802$ ; PR<sup>+</sup> patients,  $n = 525$ ; ER<sup>+</sup> PR<sup>+</sup> patients,  $n = 346$ ) the data sets GSE1456, GSE2034, GSE2990, GSE3494, GSE4922, GSE6532, GSE7390, GSE11121, GSE12093, GSE5327, GSE9195, GSE16391, GSE12276, GSE2603, GSE17705, GSE21653, GSE16446, GSE17907, GSE19615, GSE20685, GSE20711, GSE26971, GSE31448, GSE31519, E-MTAB-365, GSE20194 and GSE20271 were analysed and Kaplan–Meier plots were generated

using <http://kmplot.com> (ref. 24). The patient samples were grouped as either high or low expressor for the genes of interest based on the expression level of the selected gene, and the upper tertile were used as the cut-off and median is computed over the entire data set.

**Chromatin immunoprecipitation sequencing.** ChIP assay was carried out as described previously<sup>31</sup>. The ChIP-seq results have been deposited to Gene Expression Omnibus (GEO) under accession number GSE61198. We performed ChIP-seq experiments using either an anti-Slug or anti-Snail antibody. In the Slug-high pBL1G cells, as anticipated, we could only recover chromatin fragments from Slug ChIP but not Snail ChIP. Conversely, from the Snail<sup>high</sup> pBL3G cells, chromatin fragments were only recoverable from Snail ChIP but not Slug ChIP, demonstrating the specificities of these antibodies. Slug- or Snail-bound DNA sites were determined with Model-based Analysis of ChIP-Seq (MACS) algorithm, and bound target genes were defined as containing Slug or Snail occupancy within 5 kb upstream and downstream relative to the transcription start site for each RefSeq transcript<sup>35</sup>. Fold enrichment of each MACS peak was calculated against the whole-cell extract. Enriched motifs were identified using 1,000 nucleotides centred at the peak summit of the top 1,000 Slug and Snail peaks (ranked by MACS peak scores). The sequences were processed through MemeChIP (<http://meme.nbcr.net/meme/cgi-bin/meme-chip.cgi>) using default settings.

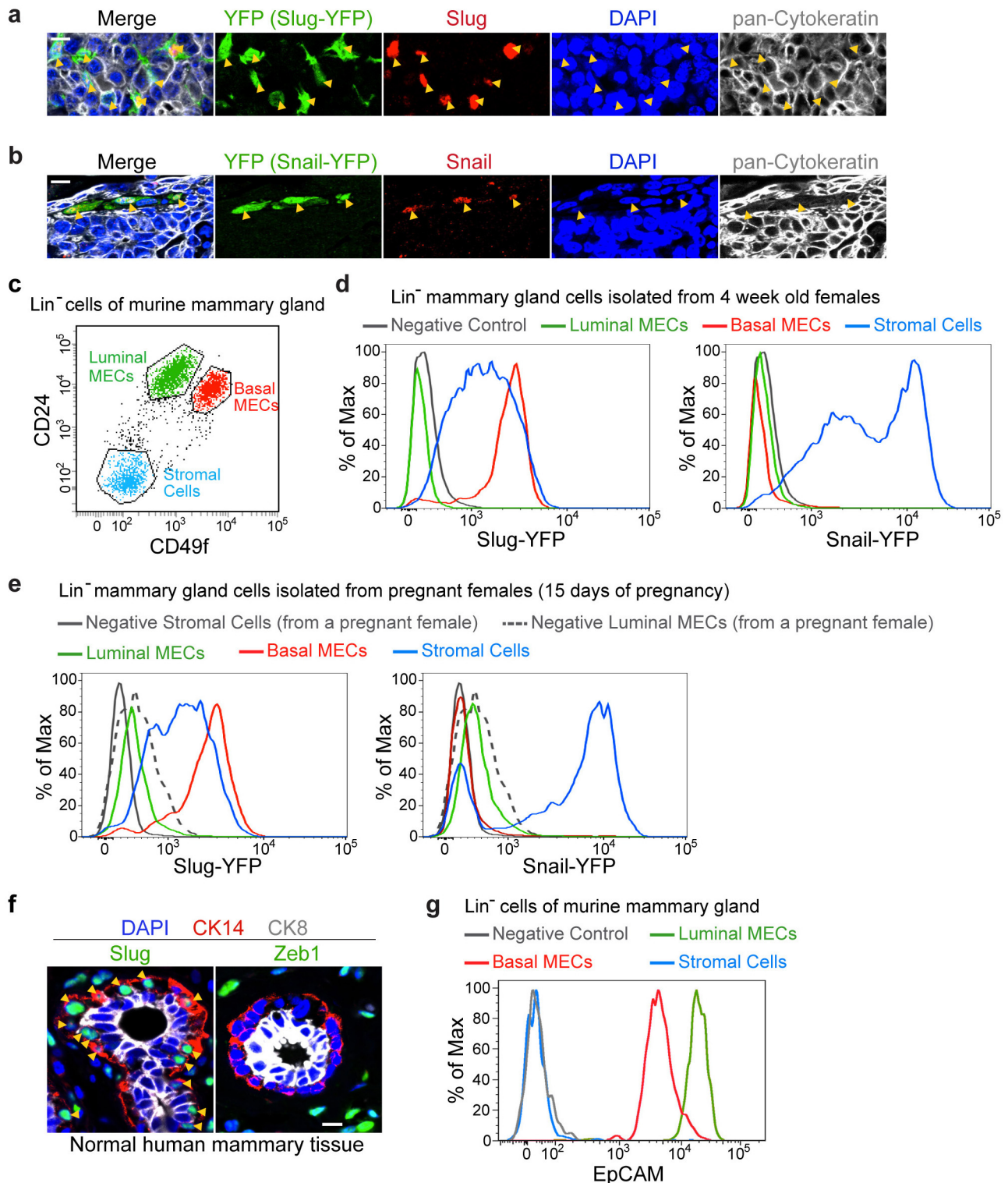
Gene set enrichment analyses (GSEA) were performed with the GSEA platform of the Broad Institute (<http://www.broadinstitute.org/gsea/index.jsp>). Slug-bound genes and Snail-bound genes were ranked according to the fold enrichment of the correspondent MACS peaks.

For SLUG and SNAIL ChIP-qPCR in MDA-MB-231 cells, ChIP-enriched DNA was analysed by real-time PCR using the ABI PRISM 7900 sequence detection system and SYBR green master mix. Relative occupancy values were calculated by determining the apparent immunoprecipitation efficiency (ratios of the amount of immunoprecipitated DNA to that of the input sample) and normalized to the level observed. The primers used for the real-time PCR are *ZEB1* target locus forward ACAAGCGAGAGGATCATGGCG, reverse CACTCACCGTTATTGCGCCG; *ZEB1* control locus forward TAATAATGGGCGGCAACGGC, reverse AGGAA CCAAAGCGAGCCCCCT.

**Statistical analysis.** Statistical analyses were carried out by two-tailed Student's *t*-test unless otherwise specified. No statistical methods were used to predetermine sample size.

31. Tam, W. L. *et al.* Protein kinase C $\alpha$  is a central signaling node and therapeutic target for breast cancer stem cells. *Cancer Cell* **24**, 347–364 (2013).
32. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).
33. Shibue, T., Brooks, M. W. & Weinberg, R. A. An integrin-linked machinery of cytoskeletal regulation that enables experimental tumor initiation and metastatic colonization. *Cancer Cell* **24**, 481–498 (2013).
34. Gupta, P. B. *et al.* The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. *Nature Genet.* **37**, 1047–1054 (2005).
35. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

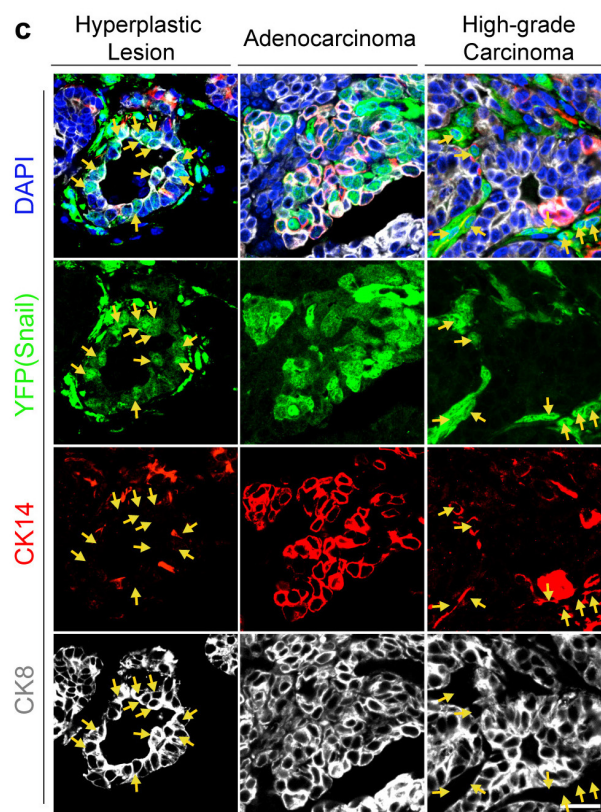
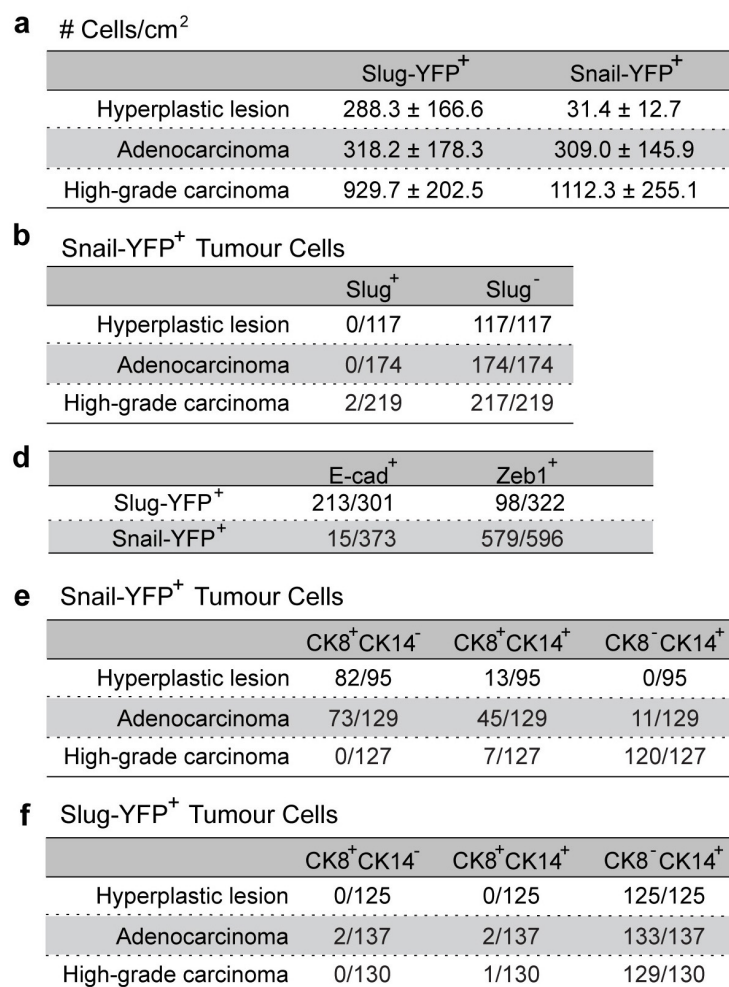




**Extended Data Figure 1 | Slug expression is associated with a partial EMT phenotype in normal MECs.** **a**, Validation of the Slug-YFP knock-in reporter. Mammary tumour sections from *Slug<sup>YFP/+</sup>;MMTV-PyMT* female mice were stained for YFP (green), Slug (red), cyokeratin (grey) and DAPI (blue). **b**, Validation of the Snail-YFP knock-in reporter. Mammary tumour sections from *Snail<sup>YFP/+</sup>;MMTV-PyMT* female mice were stained for YFP (green), Snail (red), cyokeratin (grey) and DAPI (blue). **c**, Lin<sup>-</sup> cells of normal mammary glands were separated into luminal MECs, basal MECs and stromal fibroblasts using CD24 and CD49f cell-surface markers. **d**, **e**, Representative FACS histogram showing relative expression levels of Slug-YFP and Snail-YFP

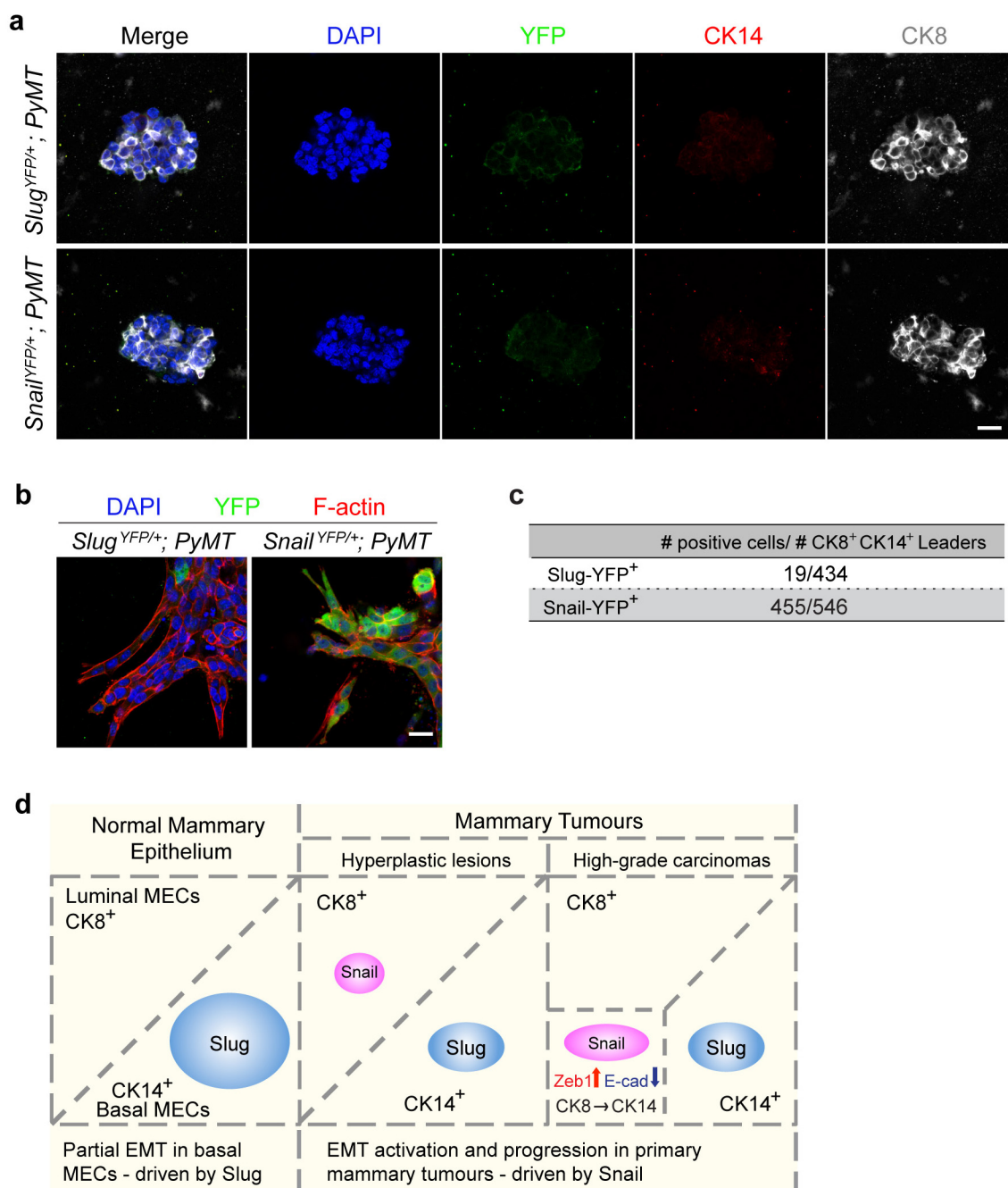
reporters in the indicated cell populations in mammary glands during puberty (**d**) and during pregnancy (**e**). Note that luminal MECs from pregnant females exhibit higher levels of autofluorescence signals (grey dashed line in panel **e**). **f**, Normal human mammary tissue sections were stained for Slug or Zeb1 (green), CK14 (red), CK8 (grey), and DAPI (blue). Arrowheads indicate *Slug<sup>+</sup>CK14<sup>+</sup>* cells. **g**, Representative FACS histogram showing expression level of the epithelial cell-surface marker EpCAM in the indicated populations of the normal mammary gland. Panels **d**, **e**, **g**, are representative of three independent experiments. All scale bars indicate 10  $\mu$ m.





**Extended Data Figure 2 | Differential expression of Snail and Slug in mammary tumours.** **a, b**, Quantifications of the frequencies of Slug-YFP<sup>+</sup> and Snail-YFP<sup>+</sup> tumour cells (**a**) and quantifications of Slug versus Snail expression (**b**) at different stages of mammary tumour development by immunofluorescence staining. For each stage, tumours from six animals were analysed for the quantifications. **c**, Individual channels of the stained image

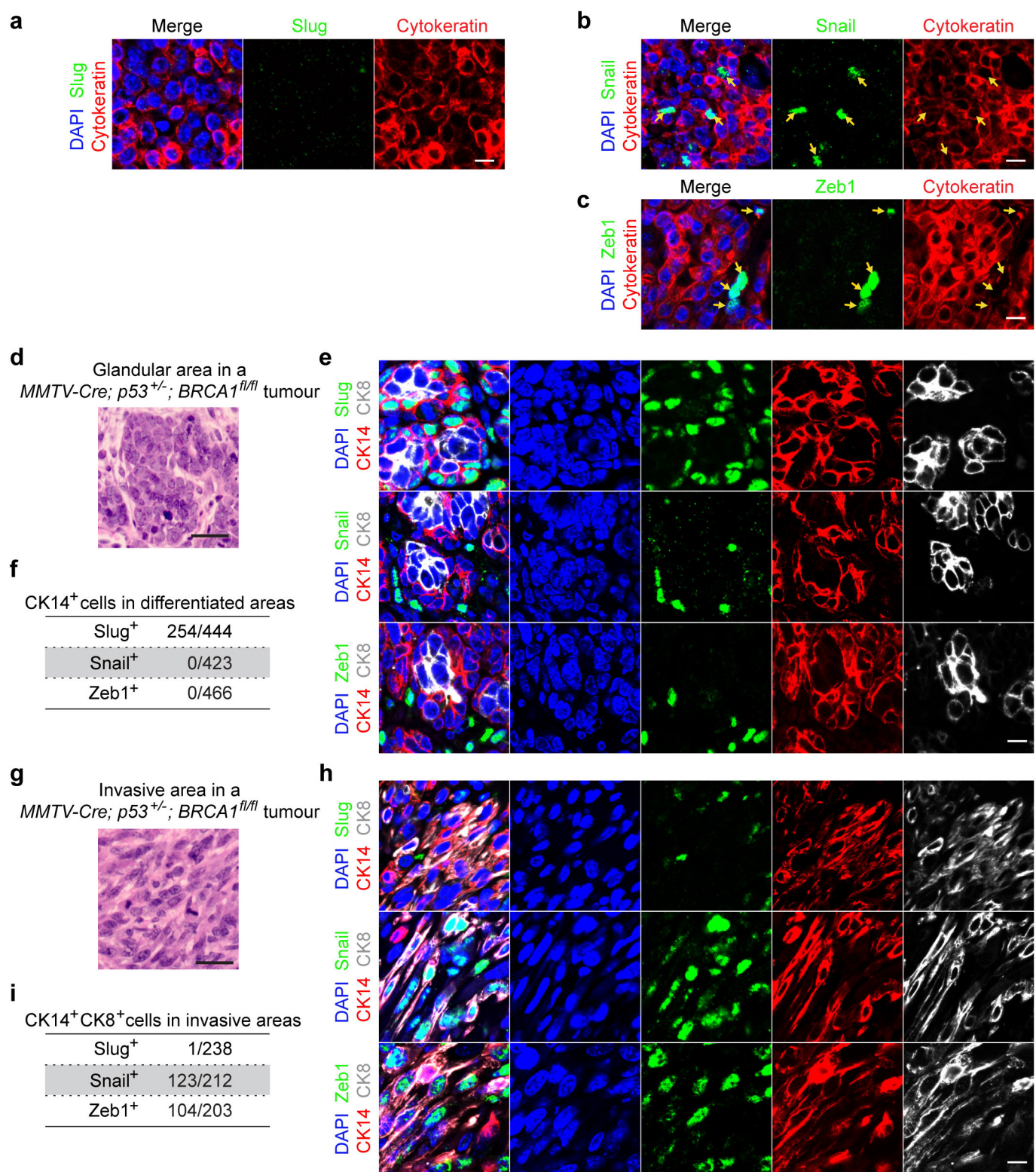
in Fig. 2e. **d**, Quantifications of E-cadherin and Zeb1 positivity (high-grade carcinomas from six animals were quantified). **e, f**, Quantification of CK8 and CK14 expression profile of Snail-YFP-positive and Slug-YFP-positive tumour cells. For each stage, tumours from six animals were analysed for the quantifications. **b, d-f**, *n* is number of cells.



**Extended Data Figure 3 | Snail activation is associated with invasive changes in mammary tumour cells *ex vivo*.** **a**, Freshly isolated tumour organoids stained for YFP (green), CK14 (red), CK8 (grey) and DAPI (blue). Note that only background staining was detected for YFP and CK14. Scale bar, 20  $\mu$ m. **b**, Tumour organoids from animals of the indicated genotypes were cultured in type I collagen gel for 48 h and stained for YFP (green), phalloidin

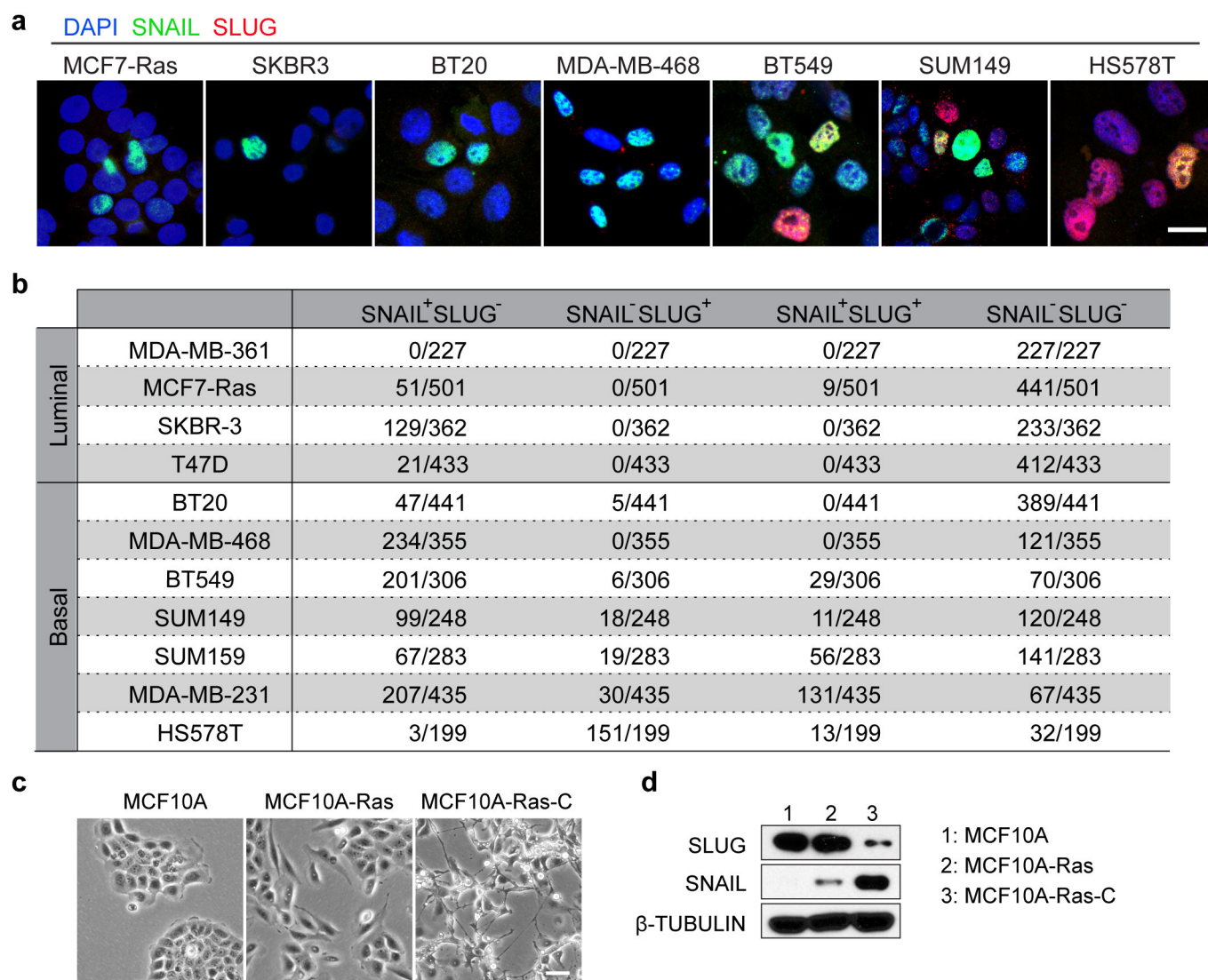
(red) and DAPI (blue). Scale bar, 10  $\mu$ m. **c**, Frequency of CK8<sup>+</sup>CK14<sup>+</sup> leader cells expressing Slug-YFP and Snail-YFP (*n*, number of cells). Tumour organoids from five different animals were analysed for each genotype. **d**, Schematic diagram summarizing expression patterns of Snail and Slug in the normal mammary gland and at different stages of mammary tumour development in the MMTV-PyMT model.





**Extended Data Figure 4 | Differential expression of Snail and Slug in MMTV-Neu and BRCA-1/p53-minus models of mammary tumours.** **a–c**, Representative immunofluorescence images of sections of aggressive MMTV-Neu tumours stained for DAPI (blue), cytokeratin (red) and Slug (green, **a**), Snail (green, **b**) or Zeb1 (green, **c**). Scale bar, 10  $\mu$ m. **d**, H&E staining showing representative histology of differentiated area in *MMTV-cre;p53<sup>+/-</sup>; BRCA1<sup>fl/fl</sup>* tumours. Scale bar, 50  $\mu$ m. **e**, Representative immunofluorescence images of the differentiated areas in *MMTV-cre;p53<sup>+/-</sup>; BRCA1<sup>fl/fl</sup>* tumours

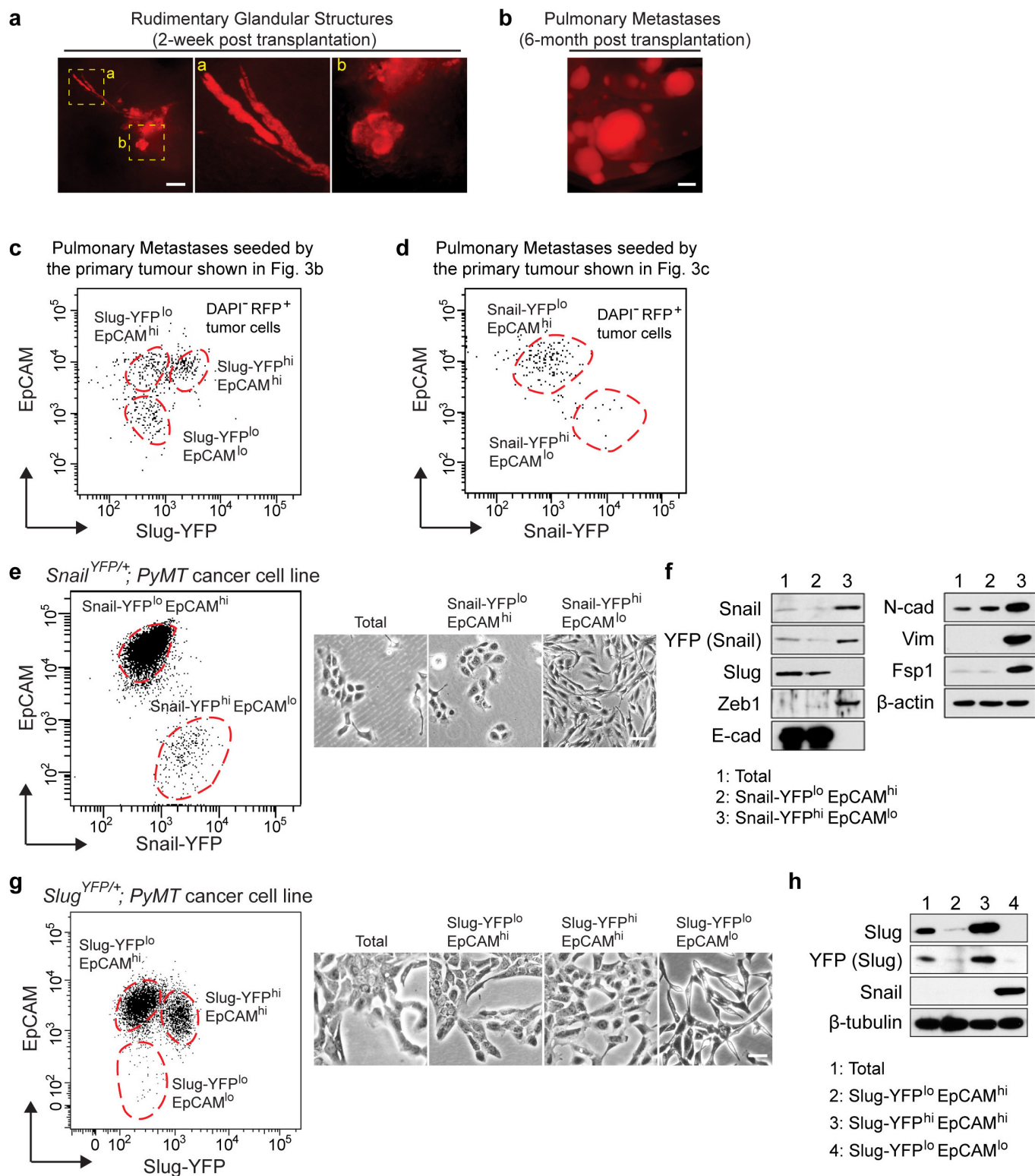
stained for the indicated proteins. Five tumours were analysed, and quantifications are shown in **f** (*n*, number of cells). Scale bar, 10  $\mu$ m. **g**, H&E staining showing representative histology of differentiated area in *MMTV-cre;p53<sup>+/-</sup>; BRCA1<sup>fl/fl</sup>* tumours. Scale bar, 50  $\mu$ m. **h**, Representative immunofluorescence images of the invasive areas in *MMTV-cre;p53<sup>+/-</sup>; BRCA1<sup>fl/fl</sup>* tumours stained for the indicated proteins. Five tumours were analysed, and quantifications are shown in **i** (*n*, number of cells).



**Extended Data Figure 5 | Differential expression of Snail and Slug in human breast cancer cell lines.** **a**, Representative immunofluorescence images of indicated human breast cancer cell lines stained for DAPI (blue), SNAIL (green) and SLUG (red). Scale bar, 10  $\mu$ m. **b**, Quantification of SLUG versus SNAIL expression in indicated human breast cancer cell lines ( $n$ , number of cells). Five fields were counted for each cell line. **c**, Representative image

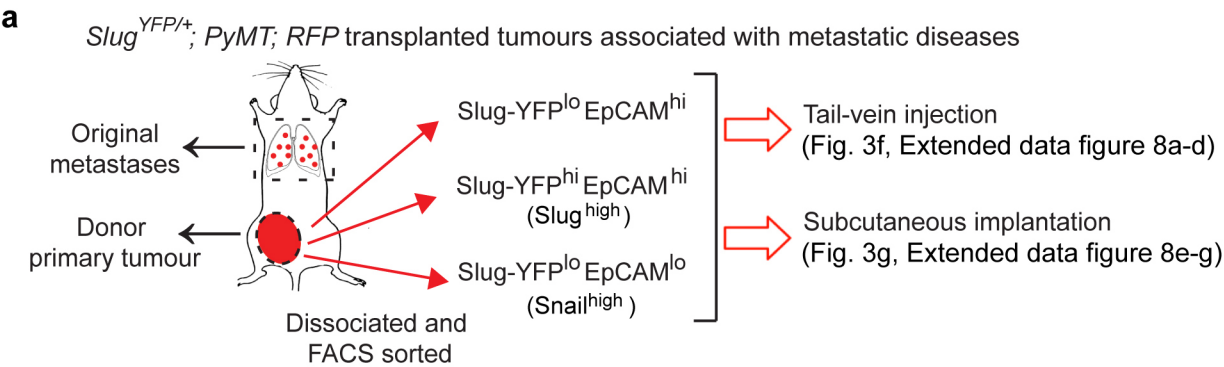
showing the morphologies of the series of MCF10A cell lines in culture. Scale bar, 50  $\mu$ m. **d**, Western blot showing expression of SLUG and SNAIL in the indicated MCF10A cell lines. Panels **a**–**d** represent two independent experiments. Uncropped western blots are available in Supplementary Information.





**Extended Data Figure 6 | Isolation of tumour cell subpopulations with differential Snail and Slug expression by FACS.** **a, b**, Representative whole-mount images showing tumour progression in the transplantation model of mammary tumours illustrated in Fig. 3a. The implanted cells initially formed rudimentary gland-like structures (**a**) and eventually progressed to become high-grade carcinomas that spontaneously metastasize to the lungs. The RFP marker allows detection of pulmonary metastases as shown in **b**. Scale bars, 500  $\mu$ m. Images represent five independent experiments. **c, d**, FACS profiles of RFP<sup>+</sup> tumour cells in the pulmonary metastases corresponding to the primary tumours shown in Fig. 3b, c. Major populations are outlined with dashed

circles. **e**, *Snail*<sup>YFP/+</sup>; *MMTV-PyMT* tumour cells were separated into indicated populations by FACS. The morphologies of the unfractionated cells and the purified populations are shown. Scale bar, 50  $\mu$ m. **f**, Western blots showing expression of EMT markers in the indicated cell populations. **g**, *Slug*<sup>YFP/+</sup>; *MMTV-PyMT* tumour cells were separated into indicated populations by FACS. The morphologies of the unfractionated cells and the purified populations are shown. Scale bar, 50  $\mu$ m. **h**, Western blots showing expression of Slug, YFP and Snail in the indicated cell populations. Uncropped western blots are available in Supplementary Information. **e–h**, Data represent three independent experiments.



**b** Tumour incidence of *Snail*<sup>YFP/+</sup>; *PyMT* cancer cell subpopulations (implanted with 10% Matrigel)

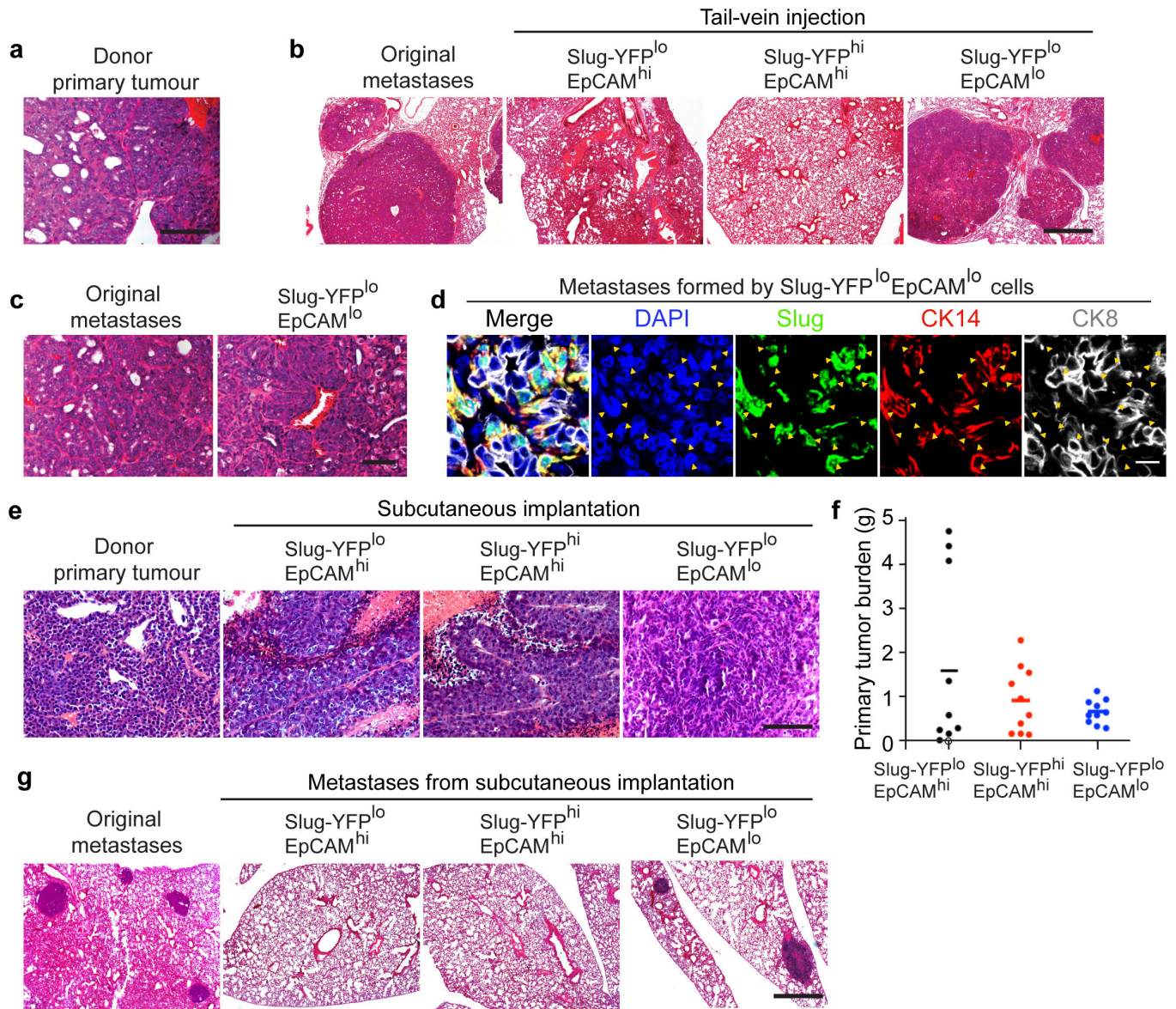
	1x10 <sup>5</sup>	1x10 <sup>4</sup>	1x10 <sup>3</sup>	1x10 <sup>2</sup>	Estimated TIC Frequency (95% Confidence Interval)
Snail-YFP <sup>hi</sup> EpCAM <sup>lo</sup>	6/6	6/6	6/6	3/6	1/143 (1/428~1/47)
Snail-YFP <sup>lo</sup> EpCAM <sup>hi</sup>	5/6	2/6	0/6	0/6	1/46,014 (1/111,052~1/19,065)

**c** Tumour incidence of *Slug*<sup>YFP/+</sup>; *PyMT* cancer cell subpopulations (implanted with 10% Matrigel)

	1x10 <sup>6</sup>	1x10 <sup>5</sup>	1x10 <sup>4</sup>	1x10 <sup>3</sup>	1x10 <sup>2</sup>	Estimated TIC Frequency (95% Confidence Interval)
Slug-YFP <sup>lo</sup> EpCAM <sup>hi</sup>	6/6	4/6	2/6	0/6	0/6	1/68,189 (1/167,066 ~ 1/27,832)
Slug-YFP <sup>hi</sup> EpCAM <sup>hi</sup>	6/6	5/6	0/6	0/6	0/6	1/72,118 (1/177,908 ~ 1/29,234)
Slug-YFP <sup>lo</sup> EpCAM <sup>lo</sup>	N.D.	6/6	6/6	6/6	4/6	1/91 (1/254 ~ 1/33)

**Extended Data Figure 7 | Fractionation of primary mammary tumours.**  
**a**, Experimental scheme for Fig. 3f, g and Extended Data Fig. 8a–g. **b, c**, Tumour cell subpopulations from *Snail*<sup>YFP/+</sup>; *MMTV-PyMT* tumour cell line (**b**) and *Slug*<sup>YFP/+</sup>; *MMTV-PyMT* tumour cell line (**c**) were injected subcutaneously at

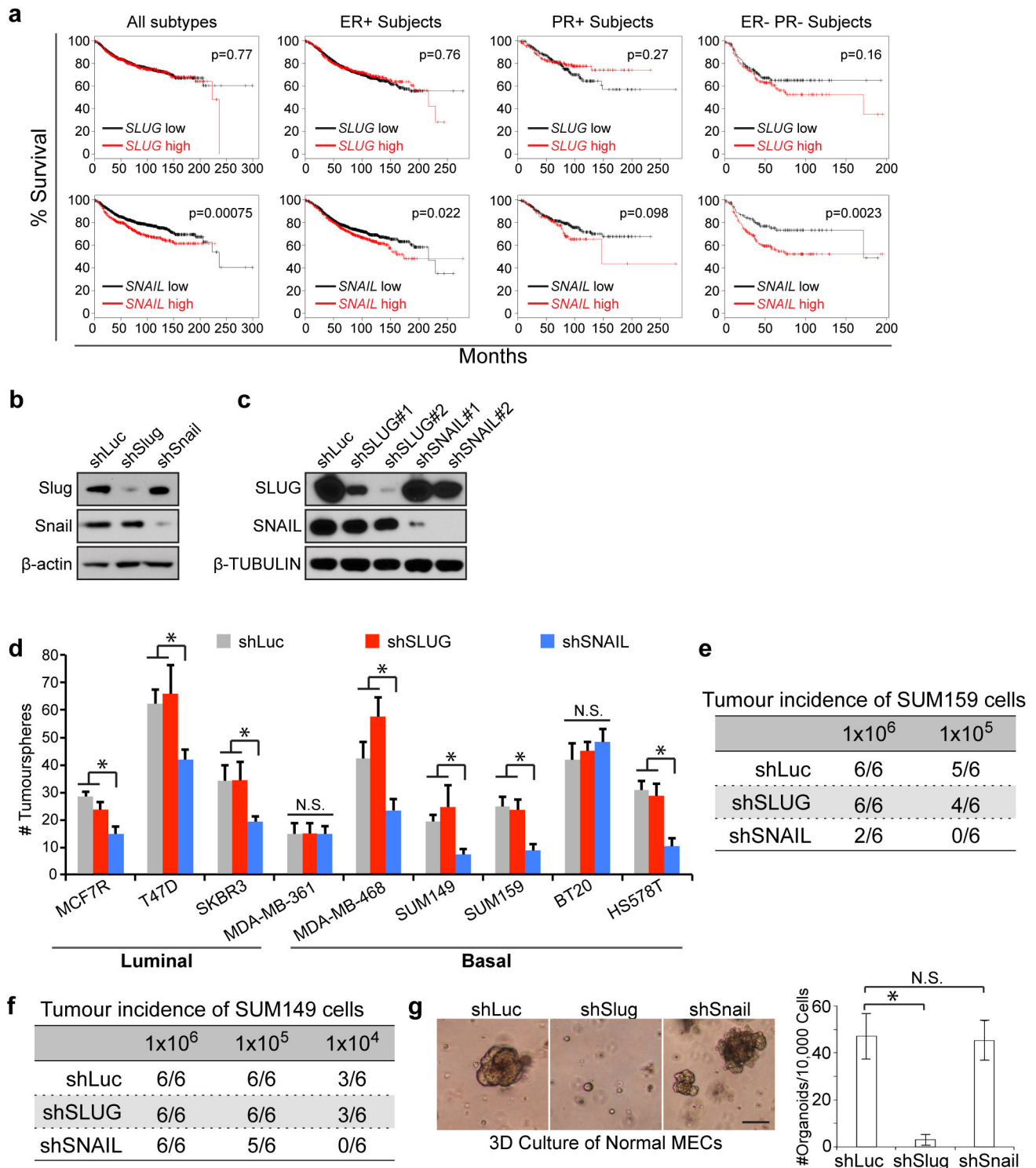
limiting dilutions to score primary tumour formation. Tumour-initiation cell frequencies were evaluated by ELDA. **b, c**, Tumour initiation was scored and presented as (number of tumour incidences/number of injections).



**Extended Data Figure 8 | Breast TICs express Snail.** **a**, H&E staining showing the histology of the donor primary tumour where cells used in Fig. 3f were isolated from. Scale bar, 200  $\mu$ m. **b**, The original pulmonary metastases spawned by the primary tumour (left panel), and pulmonary metastases formed by the indicated tumour cell populations following tail-vein injection. Scale bar, 500  $\mu$ m. **c**, Higher magnification images of H&E-stained lung sections showing histology of the original pulmonary metastases in the donor animal (left panel), and pulmonary metastases formed by the Slug-YFP<sup>lo</sup>EpCAM<sup>lo</sup> tumour cells following tail-vein injection. Scale bar, 200  $\mu$ m. **d**, Representative immunofluorescence staining image of sections of pulmonary metastases formed by the Slug-YFP<sup>lo</sup>EpCAM<sup>lo</sup> tumour cells were stained for DAPI (blue), Slug (green), CK14 (red) and CK8 (grey). Arrowheads indicate Slug-positive cells. Scale bar, 20  $\mu$ m. Images represent

four independent experiments. **e**, H&E staining of the donor primary tumour where cells used in Fig. 3g were isolated from (left panel) and H&E staining of primary tumours formed by the indicated populations following subcutaneous implantation (with 25% Matrigel). Scale bar, 200  $\mu$ m. **f**, Primary tumour burdens formed by the indicated populations after subcutaneous implantation (for EpCAM<sup>low</sup>Slug<sup>low</sup> cells  $1 \times 10^4$  cells were injected, for the other two groups  $1 \times 10^5$  cells were injected). Primary tumours and lungs were analysed 10 weeks post injection ( $n = 10$  sites of injection for each group). Open circle indicates failure of tumour initiation. Source Data is associated with this figure. **g**, H&E staining of lung sections showing metastatic outgrowths spawned by the indicated cell populations following subcutaneous implantation. Scale bar, 500  $\mu$ m.

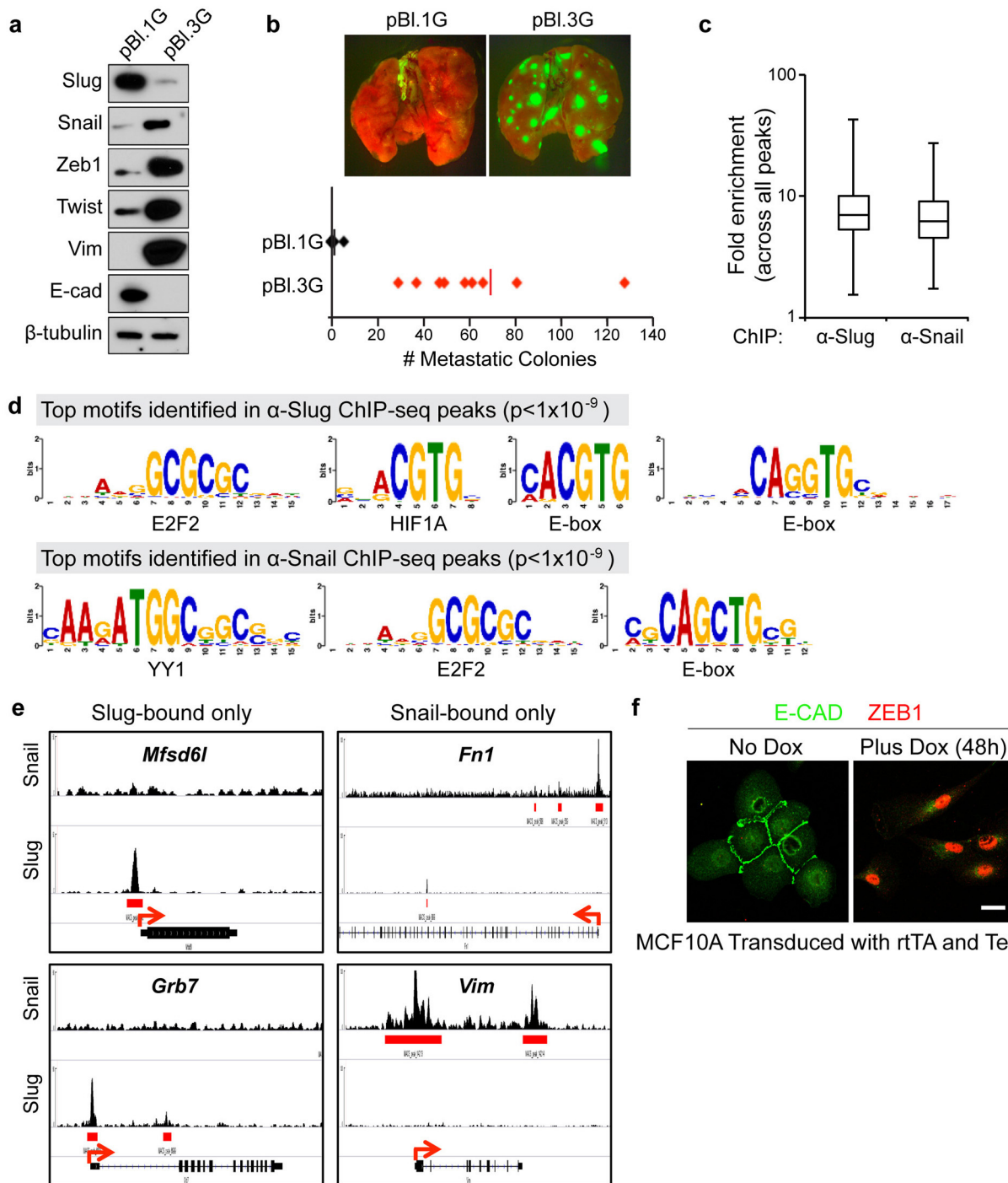




**Extended Data Figure 9 | Snail and Slug are differentially employed by normal MaSCs and breast TICs.** **a**, Kaplan-Meier plots showing survival of patients with the indicated subtypes of breast cancers. Patient groups were separated based on *SLUG* (top row) or *SNAIL* (bottom row) mRNA expression. **b**, Western blot confirming Slug and Snail knockdown in established PyMT tumour cell line transduced with the indicated shRNA expression vectors. The shLuciferase (shLuc) shRNA was used as a control. **c**, Western blot confirming *SLUG* and *SNAIL* knockdown in MDA-MB-231 cells transduced with the indicated shRNA expression vectors. shLuc was used as a control. Uncropped western blots are available in Supplementary Information. **d**, Tumour-sphere formation efficiencies (no. tumour spheres/1,000 cells for MDA-MB-361 cells,

and no. tumour spheres/200 cells for all the other cell lines) of the indicated human breast cancer cells transduced with shSLUG#2, shSNAIL#2 and the shLuc control (mean  $\pm$  s.d.,  $n = 5$  technical replicates per group). Data represent two independent experiments. **e**, **f**, SUM159 (**e**) and SUM149 (**f**) cells transduced with the indicated shRNAs were injected subcutaneously at limiting dilutions to score primary tumour formation. Tumour initiation was scored and presented as (no. of tumour incidences/no. of injections). Data represent two independent experiments. **g**, The organoid forming efficiencies of normal MECs transduced with the indicated shRNA expression vectors (mean  $\pm$  s.d.,  $n = 6$  technical replicates per group, \* $P < 0.001$ , NS, not significant.). Scale bar, 100  $\mu$ m. Data represent three independent experiments.





#### Extended Data Figure 10 | Slug and Snail occupy different genomic regions.

**a**, Western blots showing expression of EMT-TFs and EMT markers in the PyMT tumour cell lines used for the ChIP-seq analyses. Uncropped western blots are available in Supplementary Information. Data represent three independent experiments. **b**, Pulmonary metastases formed by 100,000 cells of the indicated cell lines following tail-vein injection ( $n = 9$  animals per group). Source Data is associated with this figure. **c**, Box plot showing distributions of fold enrichment of all peaks identified in Snail ChIP and Slug ChIP. Horizontal bar indicates the median and whiskers indicate the top and bottom

tertiles. **d**, Sample top motifs enriched around the summits of the anti-Snail and anti-Slug ChIP peaks. **e**, Sample ChIP-seq signals for Slug and Snail are shown. Left column shows promoters bound by Slug only. Right column shows promoters bound by Snail only. Arrows indicate the directions of transcription. **f**, MCF10A human mammary epithelial cells were transduced with rtTA and SNAIL driven by a tet-on promoter, untreated (left panel) or treated with  $2 \mu\text{g ml}^{-1}$  doxycycline (dox) for 48 h (right panel), and stained for E-cadherin (green) and ZEB1 (red). Scale bar, 20  $\mu\text{m}$ . Data represent five independent experiments.

# A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity

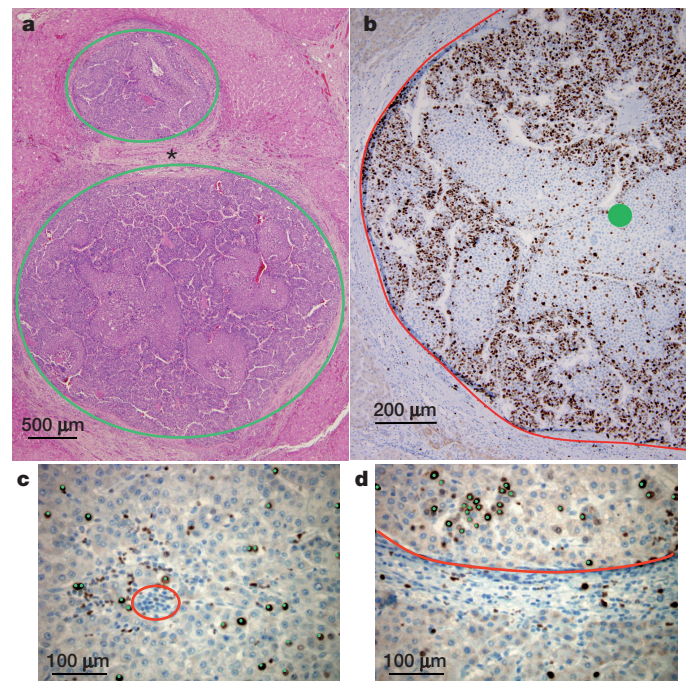
Bartłomiej Waclaw<sup>1</sup>, Ivana Bozic<sup>2,3</sup>, Meredith E. Pittman<sup>4</sup>, Ralph H. Hruban<sup>4</sup>, Bert Vogelstein<sup>4,5</sup> & Martin A. Nowak<sup>2,3,6</sup>

Most cancers in humans are large, measuring centimetres in diameter, and composed of many billions of cells<sup>1</sup>. An equivalent mass of normal cells would be highly heterogeneous as a result of the mutations that occur during each cell division. What is remarkable about cancers is that virtually every neoplastic cell within a large tumour often contains the same core set of genetic alterations, with heterogeneity confined to mutations that emerge late during tumour growth<sup>2–5</sup>. How such alterations expand within the spatially constrained three-dimensional architecture of a tumour, and come to dominate a large, pre-existing lesion, has been unclear. Here we describe a model for tumour evolution that shows how short-range dispersal and cell turnover can account for rapid cell mixing inside the tumour. We show that even a small selective advantage of a single cell within a large tumour allows the descendants of that cell to replace the precursor mass in a clinically relevant time frame. We also demonstrate that the same mechanisms can be responsible for the rapid onset of resistance to chemotherapy. Our model not only provides insights into spatial and temporal aspects of tumour growth, but also suggests that targeting short-range cellular migratory activity could have marked effects on tumour growth rates.

Tumour growth is initiated when a single cell acquires genetic or epigenetic alterations that change the net growth rate of the cell (birth minus death), and enable its progeny to outgrow surrounding cells. As these small lesions grow, the cells acquire additional alterations that cause them to multiply even faster and to change their metabolism to survive better the harsh conditions and nutrient deprivation. This progression eventually leads to a malignant tumour that can invade surrounding tissues and spread to other organs. Typical solid tumours contain about 30–70 clonal amino-acid-changing mutations that have accumulated during this multi-stage progression<sup>1</sup>. Most of these mutations are believed to be passengers that do not affect growth, and only ~5–10% are drivers that provide cells with a small selective growth advantage. Nevertheless, a major fraction of the mutations, particularly the drivers, are present in 30–100% of neoplastic cells in the primary tumour, as well as in metastatic lesions derived from it<sup>2,5</sup>.

Most attempts at explaining the genetic make-up of tumours assume well-mixed populations of cells and do not incorporate spatial constraints<sup>6–10</sup>. Several models of the genetic evolution of expanding tumours have been developed in the past<sup>11–14</sup>, but they assume either very few mutations<sup>11,12</sup> or one- or two-dimensional growth<sup>13,14</sup>. Conversely, models that incorporate spatial limitations have been developed to help to understand processes such as tumour metabolism<sup>15</sup>, angiogenesis<sup>16,17</sup> and cell migration<sup>12</sup>, but these models ignore genetics. Here, we formulate a model that combines spatial growth and genetic evolution, and use the model to describe the growth of primary tumours and metastases, as well as the development of resistance to therapeutic agents.

We first model the expansion of a metastatic lesion derived from a cancer cell that has escaped its primary site (for example, breast or colorectal epithelium) and travelled through the circulation until it lodged at a distant site (for example, lung or liver). The cell initiating the metastatic lesion is assumed to have all the driver gene mutations needed to expand. Motivated by histopathological images (Fig. 1a), we model the lesion as a conglomerate of ‘balls’ of cells (see Methods and Extended Data Fig. 1). Cells occupy sites in a regular three-dimensional lattice (Extended Data Fig. 2a, b). Cells replicate stochastically with rates proportional to the number of surrounding empty sites



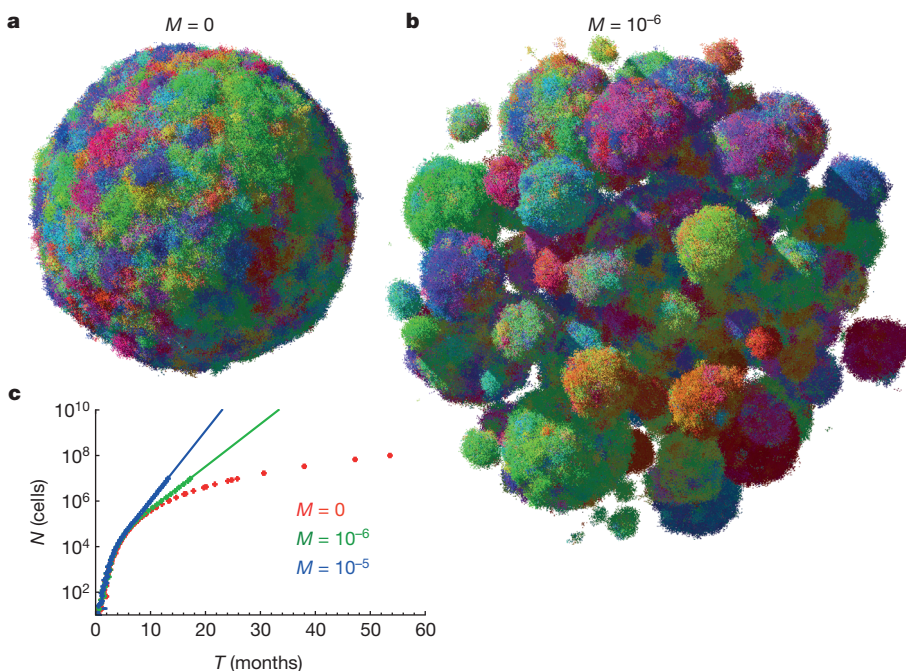
**Figure 1 | Structure of solid neoplasms.** **a**, Hepatocellular carcinoma composed of balls of cells (circled in green) separated by non-neoplastic tissue (asterisk). **b**, Adjacent section of the bottom tumour in **a** immunolabelled with the proliferation marker Ki67. The edge of the tumour is delineated in red; the centre is marked with a green circle. Proliferation is decreased in the centre when compared to the edge of the neoplasm. **c**, **d**, Higher magnification of the centre (**c**) and the edge (**d**) with each proliferating neoplastic cell marked by a green dot. The blue nuclei without green dots are non-proliferating. The red circle in **c** demonstrates an example of cells (inflammatory cells) that were not included in the count of neoplastic cells. The neoplastic tissue in **d** is above the red line; non-neoplastic (normal liver) is below the red line. Comparison of **c** with **d** shows that proliferation of neoplastic cells is decreased in the centre as compared to the edge of the lesion (quantified in Extended Data Table 1).

<sup>1</sup>School of Physics and Astronomy, University of Edinburgh, JCMB, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK. <sup>2</sup>Program for Evolutionary Dynamics, Harvard University, One Brattle Square, Cambridge, Massachusetts 02138, USA. <sup>3</sup>Department of Mathematics, Harvard University, One Oxford Street, Cambridge, Massachusetts 02138, USA. <sup>4</sup>The Sol Goldman Pancreatic Cancer Research Center, Department of Pathology, Johns Hopkins University School of Medicine, 401 North Broadway, Weinberg 2242, Baltimore, Maryland 21231, USA. <sup>5</sup>Ludwig Center and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, 1650 Orleans Street, Baltimore, Maryland 21287, USA. <sup>6</sup>Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.



(non-neoplastic cells or extracellular matrix), hence replication is faster at the edge of the tumour. This is supported by experimental data (Fig. 1b–d and Extended Data Table 1). A cell with no cancer cell neighbours replicates at the maximal rate of  $b = \ln(2) = 0.69 \text{ days}^{-1}$ , in which  $b$  denotes the initial birth rate, equivalent to 24 h cell-doubling time, and a cell that is completely surrounded by other cancer cells does not replicate. Cells can also mutate, but we assume all mutations are passengers (they do not confer fitness advantages). After replication, a cell moves with a small probability ( $M$ ) to a nearby place close to the surface of the lesion and creates a new lesion. This ‘sprouting’ of initial lesions could be due to short-range migration after an epithelial-to-mesenchymal transition<sup>18</sup> and consecutive reversion to a non-motile phenotype. Alternatively, it could be the result of another process such as angiogenesis (Methods), through which the tumour gains better access to nutrients. The same model governs the evolution of larger metastatic lesions that have already developed extensive vasculature. Cells die with a death rate ( $d$ ) independent of the number of neighbours, and are replaced by empty sites (non-neoplastic cells within the local tumour environment).

If there is little dispersal ( $M \approx 0$ ), the shape of the tumour becomes roughly spherical as it grows to a large size (Fig. 2a and Supplementary Video 2). However, even a very small amount of dispersal markedly affects the predicted shape. For  $M > 0$ , the tumour forms a conglomerate of ‘balls’ (Fig. 2b, Extended Data Fig. 2c and Supplementary Video 3), much like those observed in actual metastatic lesions, with the balls separated by islands of non-neoplastic stromal cells mixed with extracellular matrix. In addition to this remarkable change in topology, dispersal strongly affects the growth rate and doubling time of the tumour. Although the size ( $N$ ) of the tumour increases with time ( $T$ ) from initiation as  $\sim T^3$  without dispersal (Extended Data Fig. 3a, b), it grows much faster ( $\sim \exp(\text{const} \times T)$  for large  $T$ ) when  $M > 0$  (Fig. 2c). This also remains true for long-range dispersal in which  $M$  affects the probability of escape from the primary tumour into the circulation to create new lesions in distant organs (metastasis). Using plausible estimates for the rates of cell birth, death and dispersal probability, we calculate that it takes 8 years for a lesion to grow from one cell to one billion cells in the absence of dispersal ( $M = 0$ ), but less than 2 years with dispersal (Fig. 2c). The latter estimate is consistent with experimentally determined rates of metastasis growth as well as clinical experience, while the conventional model (without dispersal) is not.

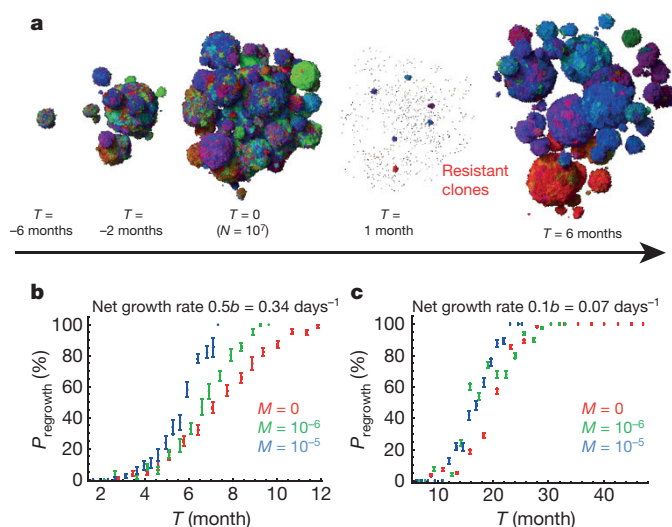


**Figure 2 | Short-range dispersal affects size, shape and growth rate of tumours.** a, b, A spherical lesion in the absence of dispersal ( $M = 0$ ) (a) and a conglomerate of lesions (b), each initiated by a cell that has migrated from a previous lesion, for low but non-zero migration ( $M = 10^{-6}$ ). Colours reflect the degree of genetic similarity; cells with similar colours have similar genetic alterations. The death rate is  $d = 0.8b$ , which corresponds to a net growth rate of  $0.2b = 0.14 \text{ days}^{-1}$ , and  $N = 10^7$  cells. c, Dispersal ( $M > 0$ ) causes the tumour to grow faster in time. Each point = 100 samples, error bars (too small to be visible) are s.e.m. Continuous lines (extrapolation) are  $6,000 \times 10^{0.43T}$  (green),  $1,000 \times 10^{0.7T}$  (blue).

Non-spatial models point to the size of a tumour as a crucial determinant of chemotherapeutic drug resistance<sup>19–21</sup>. To determine whether a spatial model would similarly predict this dependency in a clinically relevant time frame, we calculated tumour regrowth probabilities after targeted therapies. We assume that the cell that initiates the lesion is susceptible to treatment, otherwise the treatment would have no effect on the mass, and that the probability of a resistant mutation is  $10^{-7}$  (Methods); only one such mutation is needed for a regrowth.

Figure 3a shows snapshots from a simulation (Supplementary Video 1) performed before and after the administration of a typical targeted therapy at time  $T = 0$ . At first, the size of the lesion ( $\sim 3 \text{ mm}$  at  $T = 0$ ) rapidly decreases, but 1 month later resistant clones begin to proliferate and form tumours of microscopic size. Such resistant sub-clones are predicted to be nearly always present in lesions of sizes that can be visualized by clinical imaging techniques<sup>21,22</sup>. By 6 months after treatment, the lesions have regrown to their original size. The evolution of resistance is a stochastic process—some lesions shrink to zero and some regrow (Extended Data Fig. 4a). Figure 3b, c shows the probability of regrowth versus the time from the initiation of the lesion to the onset of treatment upon varying net growth rates  $b-d$  and dispersal probabilities. Regardless of growth rate, the capacity to migrate makes it more likely that regrowth will occur sooner, particularly for more aggressive cancers, that is, those which have higher net growth rates (Fig. 3b). This conclusion is in line with recent theoretical work on evolving populations of migrating cells<sup>23</sup>. If resistant mutations additionally increase the dispersal probability before or during treatment, regrowth is faster (Extended Data Fig. 4b, c).

Having shown that the predictions of the spatial model are consistent with metastatic lesion growth and regrowth times, we turn to primary tumours. In contrast to metastatic lesions, here the situation is considerably more complex because the tumour cells are continually acquiring new driver gene mutations that can endow them with fitness advantages over adjacent cells within the same tumour. Our model of a primary tumour assumes that it is initiated via a single driver gene mutation that provides a selective growth advantage over normal neighbouring cells. Each subsequent driver gene mutation reduces the death rate as  $d = b(1-s)^k$ , in which  $k$  is the number of driver mutations in the cell ( $k \geq 1$ ), and  $s$  is the average fitness advantage per driver. Almost identical results are obtained if driver gene mutations increase cell birth rather than decrease cell death, or affect both

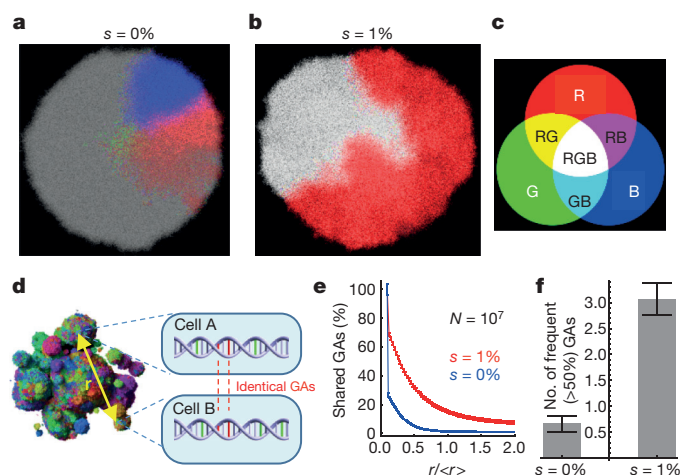


**Figure 3 | Treatment success rates depend on the net growth rate of tumours.** **a**, Time snapshots before and during therapy ( $M = 10^{-6}$ ). Resistant subpopulations that cause the tumour to regrow after treatment can be seen at  $T = 1$  month. **b**, **c**, Probability of tumour regrowth ( $P_{\text{regrowth}}$ ) as a function of time after treatment initiation, for different dispersal probabilities ( $M$ ) and net growth rates of the resistant cells. A higher net growth rate (**b**) leads to a high regrowth probability, so that 50% of tumours regrow 6 months after treatment is initiated when  $M = 10^{-5}$ . **c**, Tumours with lower net growth rates require  $>20$  months to achieve the same probability of regrowth. Number of samples = 1 to 800 per point (282 on average). Error bars are s.e.m. See Methods for details.

cell birth and cell death (Extended Data Fig. 5b); the most important parameter is the fitness gain,  $s$ , conferred by each driver mutation.

Figure 4a shows that in the absence of any new driver mutations (as for a perfectly normal cell growing in utero), clonal subpopulations would be restricted to small, localized areas. Each of these areas has at least one new genetic alteration, but none of them confers a fitness advantage (they are ‘passengers’). In an early tumour, in which the centre cell contains the initiating driver gene mutation, the same structure would be observed—as long as no new driver gene mutations have yet appeared. The occurrence of a new driver gene mutation, however, markedly alters the spatial distribution of cells. In particular, the heterogeneity observed in normal cells (Fig. 4a) is substantially reduced (Fig. 4b and Supplementary Video 5). The degree of heterogeneity can be quantified by calculating the number of genetic alterations (passengers plus drivers) shared between two cells separated by various distances (Fig. 4d–f). The genetic diversity is markedly decreased (Fig. 4e), even with relatively small fitness advantages ( $s = 1\%$ ). This also has implications for the number of genetic alterations that will be present in a macroscopic fraction (for example,  $>50\%$ ) of all cells. Figure 4f shows that this number is many times larger for  $s = 1\%$  than  $s = 0\%$ . Furthermore, our model predicts that virtually all cells within a large tumour will have at least one new driver gene mutation after 5 years of growth (Extended Data Fig. 5a). The faster the clonal expansion occurs (the larger  $s$  is), the smaller the number of passenger mutations (Extended Data Fig. 5d, e). Our results are also robust to changes to the model (Methods and Extended Data Figs 5 and 6). We stress that an important prerequisite for limiting heterogeneity is cell turnover in the tumour, because in the spatial setting cells with driver mutations can ‘percolate’ through the tumour only if they replace other cells. In the absence of cell turnover, tumours are much more heterogeneous (Extended Data Fig. 6d).

In summary, our model accounts for many facts observed clinically and experimentally. Our results are robust and many assumptions can be relaxed without qualitatively affecting the outcome (Methods and Supplementary Information). Although tumour cell migration has



**Figure 4 | Genetic diversity is strongly reduced by the emergence of driver mutations.** **a–f**, For all,  $M = 0$  and the initial net growth rate =  $0.007 \text{ days}^{-1}$  ( $d = 0.99b$ ). The three most abundant genetic alterations (GAs) have been colour-coded using red (R), green (G) and blue (B) (**c**). Each section is 80 cells thick. Combinations of the three basic colours correspond to cells having two or three of these genetic alterations. **a**, No drivers—separated, conical sectors emerge in different parts of the lesion, each corresponding to a different clone. **b**, Drivers with selective advantage  $s = 1\%$  lead to clonal expansions and many cells have all three genetic alterations (white area). **d**, Genetic diversity can be determined quantitatively by randomly sampling pairs of cells separated by distance  $r$  and counting the number of shared genetic alterations. **e**, The number of shared genetic alterations versus the normalized distance  $r/\langle r \rangle$  decreases much more slowly for the case with (red) than without (blue) driver mutations. **f**, The total number of genetic alterations present in at least 50% of all cells is much larger for  $s = 1\%$  than for  $s = 0\%$ . Number of samples = 50 per data point. Error bars are s.e.m.

historically been viewed as a feature of cancer associated with late events in tumorigenesis, such as invasion through basement membranes or vascular walls, this classical view of migration pertains to the ability of cancer cells to migrate over large distances<sup>24</sup>. Instead, our analysis reveals that even small amounts of localized cellular movement are able to markedly reshape a tumour. Moreover, we predict that the rate of tumour growth can be substantially altered by a change in dispersal rate of the cancer cells, even in the absence of any changes in doubling times or net growth rates of the cells within the tumour. Some of our predictions could be experimentally tested using new cell labelling techniques<sup>25,26</sup>. Our results could also greatly inform the interpretation of mutations in genes whose main functions seem to be related to the cytoskeleton or to cell adhesion rather than to cell birth, death, or differentiation<sup>27,28</sup>. For example, cells that have lost the expression of E-cadherin (a cell adhesion protein) are more migratory than normal cells with intact E-cadherin expression<sup>29</sup>, and loss of E-cadherin in pancreatic cancer has been associated with poorer prognosis<sup>30</sup>, in line with our predictions.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 September 2014; accepted 23 July 2015.

Published online 26 August 2015.

- Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
- Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
- Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl Acad. Sci. USA* **110**, 4009–4014 (2013).
- Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
- Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
- Gatenby, R. A. & Vincent, T. L. An evolutionary model of carcinogenesis. *Cancer Res.* **63**, 6212–6220 (2003).



7. Johnston, M. D., Edwards, C. M., Bodmer, W. F., Maini, P. K. & Chapman, S. J. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proc. Natl Acad. Sci. USA* **104**, 4008–4013 (2007).
8. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl Acad. Sci. USA* **107**, 18545–18550 (2010).
9. Beerenwinkel, N. *et al.* Genetic progression and the waiting time to cancer. *PLOS Comput. Biol.* **3**, e225 (2007).
10. Durrett, R. & Moseley, S. Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Popul. Biol.* **77**, 42–48 (2010).
11. González-García, I., Sole, R. V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl Acad. Sci. USA* **99**, 13085–13089 (2002).
12. Thalhauser, C. J., Lowengrub, J. S., Stupack, D. & Komarova, N. L. Selection in spatial stochastic models of cancer: migration as a key modulator of fitness. *Biol. Direct* **5**, 21 (2010).
13. Martens, E. A., Kostadinov, R., Maley, C. C. & Hallatschek, O. Spatial structure increases the waiting time for cancer. *New J. Phys.* **13**, 115014 (2011).
14. Anderson, A. R. A., Weaver, A. M., Cummings, P. T. & Quaranta, V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell* **127**, 905–915 (2006).
15. Kim, Y., Magdalena, A. S. & Othmer, H. G. A hybrid model for tumor spheroid growth *in vitro* I: theoretical development and early results. *Math. Models Methods Appl. Sci.* **17**, 1773–1798 (2007).
16. McDougall, S. R., Anderson, A. R. & Chaplain, M. A. Mathematical modeling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. *J. Theor. Biol.* **241**, 564–589 (2006).
17. Hawkins-Daarud, A., Rockne, R. C., Anderson, A. R. A. & Swanson, K. R. Modeling tumor-associated edema in gliomas during anti-angiogenic therapy and its impact on imageable tumor. *Front. Oncol.* **3**, 66 (2013).
18. McDonald, O. G., Wu, H., Timp, W., Doi, A. & Feinberg, A. P. Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nature Struct. Mol. Biol.* **18**, 867–874 (2011).
19. Komarova, N. L. & Wodarz, D. Drug resistance in cancer: principles of emergence and prevention. *Proc. Natl Acad. Sci. USA* **102**, 9714–9719 (2005).
20. Bozic, I., Allen, B. & Nowak, M. A. Dynamics of targeted cancer therapy. *Trends Mol. Med.* **18**, 311–316 (2012).
21. Bozic, I. & Nowak, M. A. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *Proc. Natl Acad. Sci. USA* **111**, 15964–15968 (2014).
22. Turke, A. B. *et al.* Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).
23. Komarova, N. L. Spatial interactions and cooperation can change the speed of evolution of complex phenotypes. *Proc. Natl Acad. Sci. USA* **111**, 10789–10795 (2014).
24. Talmadge, J. E. & Fidler, I. J. AACR Centennial Series: the biology of cancer metastasis: historical perspective. *Cancer Res.* **70**, 5649–5669 (2010).
25. Alcolea, M. P. *et al.* Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. *Nature Cell Biol.* **16**, 615–622 (2014).
26. Weber, K. *et al.* RGB marking facilitates multicolor clonal cell tracking. *Nature Med.* **17**, 504–509 (2011).
27. Bordeleau, F., Alcoser, T. A. & Reinhart-King, C. A. Physical biology in cancer. 5. The rocky road of metastasis: the role of cytoskeletal mechanics in cell migratory response to 3D matrix topography. *Am. J. Physiol. Cell Physiol.* **306**, C110–C120 (2014).
28. Lawson, C. D. & Burridge, K. The on-off relationship of Rho and Rac during integrin-mediated adhesion and cell migration. *Small GTPases* **5**, e27958 (2014).
29. Gall, T. M. H. & Frampton, A. E. Gene of the month: E-cadherin (CDH1). *J. Clin. Pathol.* **66**, 928–932 (2013).
30. Winter, J. M. *et al.* Absence of E-cadherin expression distinguishes noncohesive from cohesive pancreatic cancer. *Clin. Cancer Res.* **14**, 412–418 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Support from The John Templeton Foundation is gratefully acknowledged. B.W. was supported by the Leverhulme Trust Early-Career Fellowship, and the Royal Society of Edinburgh Personal Research Fellowship. I.B. was supported by Foundational Questions in Evolutionary Biology Grant RFP-12-17. M.E.P., R.H.H. and B.V. acknowledge support from The Virginia and D.K. Ludwig Fund for Cancer Research, The Lustgarten Foundation for Pancreatic Cancer Research, The Sol Goldman Center for Pancreatic Cancer Research, and NIH grants CA43460 and CA62924.

**Author Contributions** B.W., M.A.N., I.B. and B.V. designed the study. B.W. wrote the computer programs and made simulations. B.W., I.B. and M.A.N. made analytic calculations. M.E.P. and R.H.H. carried out experimental work. All authors discussed the results. The manuscript was written primarily by B.W., M.A.N., I.B. and B.V., with contributions from M.E.P. and R.H.H.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.N. ([martin.nowak@harvard.edu](mailto:martin.nowak@harvard.edu)).

## METHODS

No statistical methods were used to predetermine sample size. Experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Spatial model for tumour evolution.** Tumour modelling has a long tradition<sup>31</sup>. Many models of spatially expanding tumours were proposed in the past<sup>12–14,16,32–42</sup>, but they either assume very few<sup>32,34–37,39,41,43–47</sup> or no new mutations at all<sup>12,15,38,48,49</sup>, or one- or two-dimensional growth<sup>12,13,32,33,50–52</sup>. On the other hand, well-mixed models with several mutations<sup>6,8,53,54</sup> do not often include space, and computational models aimed at being more biologically realistic<sup>15,55–61</sup> require too much computing resources (time and memory) to simulate realistically large tumours ( $N \approx 10^9$  cells). Our model builds on the Eden lattice model<sup>62</sup> and combines spatial growth and accumulation of multiple mutations. Since we focus on the interplay of genetics, spatial expansion and short-range dispersal of cells, for simplicity we do not explicitly model metabolism<sup>14</sup>, tissue mechanics, spatial heterogeneity of tissues, different types of cells present or angiogenesis<sup>16</sup>.

A tumour is made of non-overlapping balls (microlesions) of cells. Tumour cells occupy sites of a regular 3D square lattice (Moore neighbourhood, 26 neighbours). Empty lattice sites are assumed to be either normal cells or filled with extracellular matrix and are not modelled explicitly. Each cell in the model is described by its position and a list of genetic alterations that have occurred since the initial neoplastic cell, and the information about whether a given mutation is a passenger, driver, or resistance-carrying mutation. A passenger mutation does not affect the net growth rate whereas a driver mutation increases it by disrupting tight regulation of cellular divisions and shifts the balance towards increased proliferation or decreased apoptosis. The changes can also be epigenetic and we do not distinguish between different types of alterations. We assume that each genetic alteration occurs only once ('infinite allele model'<sup>63</sup>). The average numbers of all genetic alterations, driver and resistant genetic alterations produced in a single replication event are denoted by  $\gamma$ ,  $\gamma_d$  and  $\gamma_r$ , respectively. When a cell replicates, each of the daughter cells receives  $n$  new genetic alterations of each type ( $n$  being generally different in both cells) drawn at random from the Poisson probability distribution:

$$P(n) = \frac{e^{-\gamma_x/2} (\gamma_x/2)^n}{n!} \quad (1)$$

in which  $x$  denotes the type of genetic alteration.

In model A shown in Figs 2–4, replication occurs stochastically, with rate proportional to the number of empty sites surrounding the replicating cell, and death occurs with constant rate depending only on the number of drivers. We also simulated other scenarios (models B, C and D, see below). Driver mutations increase the net growth rate (the difference between proliferation and death) either by increasing the birth rate or decreasing the death rate by a constant factor  $1 + s$ , in which  $s > 0$ .

Dispersal is modelled by moving an offspring cell to a nearby position where it starts a new microlesion (Extended Data Fig. 1a). Microlesions repel each other; a 'shoving' algorithm<sup>64,65</sup> (Extended Data Fig. 1b) ensures they do not merge.

**Code availability.** The computer code (available at <http://www2.ph.ed.ac.uk/~bwaclaw/cancer-code>) can handle up to  $1 \times 10^9$  cells, which corresponds to tumours that are clinically meaningful and can be observed by conventional medical imaging (diameter  $> 1$  cm). The algorithm is discussed in details in the Supplementary Information. It is not an exact kinetic Monte Carlo algorithm because such an algorithm would be too slow to simulate large tumours. A comparison with kinetic Monte Carlo for smaller tumours (Supplementary Information) shows that both algorithms produce consistent results.

**Model parameters.** The initial birth rate  $b = \ln(2) \approx 0.69$  days<sup>-1</sup>, which corresponds to a 24 h minimum doubling time. The initial death rate  $d = 0 \dots 0.995b$  depends on the aggressiveness of the tumour (larger values = less aggressive lesion). In simulations of targeted therapy, we assume that, before treatment,  $b = 0.69$  days<sup>-1</sup> and  $d = 0.5b = 0.35$  days<sup>-1</sup>, whereas during treatment  $b = 0.35$  days<sup>-1</sup> and  $d = 0.69$  days<sup>-1</sup>, that is, birth and death rates swap places. This rather arbitrary choice leads to the regrowth time of about 6 months, which agrees well with clinical evidence. Mutation probabilities are  $\gamma = 0.02$ ,  $\gamma_d = 4 \times 10^{-5}$ ,  $\gamma_r = 1 \times 10^{-7}$ , in line with experimental evidence and theoretical work<sup>8,66–68</sup>. Since there are no reliable data on the dispersal probability  $M$ , we have explored a range of values between  $M = 1 \times 10^{-7}$  and  $1 \times 10^{-2}$ . All parameters are summarized in Extended Data Fig. 1c, see also further discussion in Supplementary Information.

**Validity of the assumptions of the model.** Our model is deliberately oversimplified. However, many of the assumptions we make can be experimentally justified or shown not to qualitatively affect the model.

**Three-dimensional regular lattice of cells.** The 3D Moore neighbourhood was chosen because it is computationally fast and introduces relatively fewer artefacts related to lattice symmetries. Real tissues are much less regular and the number of nearest neighbours is different<sup>69</sup>. However, recent simulations of similar models of

bacterial colonies<sup>70,71</sup> show that the structure of the lattice (or the lack thereof in off-lattice models) has a marginal effect on genetic heterogeneity.

**Asynchronous cell division.** Division times of related cells remain correlated for a few generations. However, stochastic cell division implemented in our model is a good approximation for a large mass of cells and is much less computationally expensive than modelling a full cell cycle.

**Replication faster at the boundary than in the interior.** Several studies have described a higher proliferation rate at the leading edge of tumours, and this has been associated with a more aggressive clinical course<sup>72</sup>. To estimate the range of values of death rate  $d$  for our model, we used the proliferation marker Ki67. Representative formalin-fixed, paraffin-embedded tissue blocks were selected from four small chromophobe renal cell carcinomas and six small hepatocellular carcinomas by the pathologist (M.E.P.). A section of each block was immunolabelled for Ki67 using the Ventana Benchmark XT system. Around 8–12 images, depending on the size of the lesion, were acquired from each tumour. Fields were chosen at random from the leading edge and the middle of the tumour and were not necessarily 'hot spots' of proliferative activity. Using an ImageJ macro, each Ki67-positive tumour nucleus was labelled green by the pathologist, and each Ki67-negative tumour nucleus was labelled red. Other cell types (endothelium, fibroblasts and inflammatory cells) were not labelled. The proliferation rate was then calculated using previously described methods<sup>73</sup>. Statistical significance of the results was determined using a Kolmogorov–Smirnov two-sample test (significance level 0.05). The study was approved by the Institutional Review Board of the Johns Hopkins University School of Medicine. In all ten tumours, the proliferation rate at the leading edge of the tumour was greater than that at the centre by a factor of 1.25 to 6 (Extended Data Table 1). Comparing the density of proliferating cells to our model gives  $d \approx 0.5b$  (range:  $d = 0.17b \dots 0.8b$ ), which is what we assume in the simulations of aggressive lesions.

**Equal fitness of all cells in metastatic lesions.** We assume that cells in a metastatic lesion are already very fit since they contain multiple drivers. Indeed, studies of primary tumours and their matched metastases usually fail to find driver mutations present in the metastases that were not present in the primary lesions<sup>2,74</sup>, although there are notable exceptions, see, for example, refs 75 and 76. Experimental evidence in microbes<sup>77</sup> and (to a lesser extent) in eukaryotes<sup>78</sup> suggests that fitness gains due to individual mutations are largest at the beginning of an evolutionary process and that the effects of later mutations are much smaller. It remains to be seen how well these results apply to late genetic alterations in cancer<sup>79</sup> but if true, new drivers occurring in the lesion are unlikely to spread through the population before the lesion reaches a clinically relevant size.

**Dispersal.** In our model, cells detach from the lesion and attach again at a different location in the tissue. This can be viewed either as cells migrating from one place to another one, or as a more generic mechanism that allows tumour cells to get better access to nutrients by dispersing within the tissue, hence providing a growth advantage over cells that did not disperse. Some mechanisms that do not involve active motion (that is, cells becoming motile) are discussed below.

**Migration.** Cancer cells are known to undergo epithelial-to-mesenchymal transition, the origin of which is thought to be epigenetic<sup>18</sup>. This involves a cell becoming motile and moving some distance. If the cell finds the right environment, it can switch back to the non-motile phenotype and start a new lesion. Motility can be enhanced by tissue fluidization due to replication and death<sup>80</sup>. Instead of modelling the entire cycle (epithelial–mesenchymal–epithelial), we only model the final outcome (a cell has moved some distance).

**Tumour buds.** Many tumours exhibit focally invasive cell clusters, also known as tumour buds. Their proliferation rate is less than that of cells in the main tumour<sup>81</sup>. We propose that tumour buds contain cells that have not yet completed epithelial-to-mesenchymal transition and therefore they proliferate slower.

**Single versus cluster migration.** Ref. 82 found that circulating cancer cells can travel in clusters of 2–50 cells, and that such clusters can initiate metastatic foci. They report that approximately one-half of the metastatic foci they examined were initiated by single circulating cancer cells, and that circulating cancer cell clusters initiated the other half. The authors also note that the cells forming a cluster are probably neighbouring cancer cells from the primary tumour. This means that the genetic make-up of cells within a newly established lesion will be very similar, regardless of its origin (single cell versus a small cluster of cells). Therefore, the ability to travel in clusters should not affect the genetic heterogeneity or regrowth probability as compared to single-cell dispersal from our model.

**Angiogenesis.** We do not explicitly model angiogenesis for two reasons. First, most genetic alterations that can either change the growth rate or be detected experimentally must occur at early stages of tumour growth as explained before. Hence, the genetic make-up of the tumour is determined primarily by what happens before angiogenesis. Second, local dispersal from the model mimics tumour cells interspersing with the vascularized tissue and getting better access to nutrients, which is one of the outcomes of angiogenesis.

**Biomechanics of tumours.** Growth is affected by the mechanical properties of cells and the extracellular matrix. We do not explicitly include biomechanics (see, however, below), in contrast to more realistic models<sup>83,84</sup>, as this would not allow us to simulate lesions larger than about  $1 \times 10^6$  cells. Instead, we take experimentally determined values for birth and death rates, values that are affected by biomechanics, as the parameters of our model.

**Isolated balls of cells.** In our simulations, balls of cells are thought to be separated by normal, vascularized tissue which delivers nutrients to the tumour. The environment of each ball is the same, and there are no interactions between the balls other than mechanical repulsion. This represents a convenient mathematical contrivance and qualitatively recapitulates what is observed in stained sections of actual tumours (Fig. 1a). We investigated under which circumstances the balls of cancer cells would mechanically repel each other; see Extended Data Fig. 7 for a graphical summary of the results. We simulated a biomechanical, off-lattice model of normal tissue composed of ‘ducts’ lined with epithelial cells and separated by stroma (Supplementary Information, section 8). Mechanical interactions between cells were modelled using an approach similar to that described previously<sup>59,60,85</sup>, with model parameters taken from refs 59, 60, 85–88. We assumed cancer cells to be of epithelial origin, as are most cancers<sup>89</sup>. Cancer cells that invaded different areas of epithelium grew into balls that remained separated by thin slices of stroma (Supplementary Videos 8–11). This ‘encapsulation’ of tumour microlesions was possible owing to the supportive nature of stroma that is able to mechanically resist expansion of balls of cancer cells. Encapsulation is essential if the balls are to repel each other. If the tissue is ‘fluidized’ by random replication and death, the balls quickly merge (Supplementary Video 12). Another important factor are differences in mechanical properties of tumour and normal cells<sup>90</sup>; it is known that differences in cellular adhesion and stiffness promote segregation of different types of cells<sup>91,92</sup>.

In reality, microlesions within the primary tumour are less symmetric and some of them are better described as ‘protrusions’ bulging out from the main tumour tissue, owing to biomechanical instabilities; see, for example, refs 93, 94. However, stroma may still provide enough spatial separation, and the capillary network of blood vessels—either due to tumour angiogenesis or preexisting in the invaded tissue—may provide enough nutrients to the lesions so that our assumption of independently growing balls of cells remains valid. Therefore, we believe that modelling the tumour as a collection of non- or weakly-interacting microlesions is essentially correct. We also note that the existence of isolated balls is not necessary to explain our qualitative results: reduced heterogeneity and increased growth in the presence of migration. Supplementary Video 13 shows that even if the tissue is homogeneous and highly dynamical and there are no isolated balls of cells, migration leads to a considerable speedup of growth as compared to the case with no migration (Supplementary Video 14).

**Tumour geometry and heterogeneity in the absence of driver mutations.** Supplementary Videos 2 and 3 illustrate the process of growth of a tumour with maximally  $N = 10^7$  cells, for  $M = 0$  and  $M = 10^{-6}$ , respectively, and for  $d = 0.5$ . Extended Data Fig. 2 shows snapshots from a single simulation for  $M = 0$ ,  $N \sim 10^3$ , and  $d = 0$  (no death, Extended Data Fig. 2a) and  $d = 0.9$  (Extended Data Fig. 2b). In the latter case, cells are separated by empty sites (normal cells/extracellular matrix). Extended Data Fig. 2c shows that the tumour is almost spherically symmetric for  $M = 0$ . The symmetry is lost for small but non-zero  $M$ , and restored for larger  $M$  when the balls become smaller and their number increases. Extended Data Fig. 2c also shows that metastatic tumours contain many clonal sectors with passenger mutations. Extended Data Fig. 8a shows that the fraction  $G(r)$  of genetic alterations that are the same in two randomly sampled cells (Fig. 4) separated by distance  $r$  quickly decreases with  $r$ , indicating increased genetic heterogeneity owing to passenger mutations.

**Targeted therapy of metastatic lesions.** Models of cancer treatment<sup>21,95–100</sup> often assume either no spatial structure or do not model the emergence of resistance. We assume that the cell that initiated the lesion was sensitive to treatment but its progeny may become resistant. Before the therapy commences, all cells have the same birth and death rates, but after the treatment resistant cells continue to proliferate with the same rate, whereas susceptible cells are assigned different rates as described above. Resistant cells can emerge before and during the therapy. The death rate of sensitive cells during treatment is greater than the birth rate, or the tumours would not be sensitive to the drug. For example, in Fig. 3 treatment increases the death rate and decreases the growth rate of susceptible cells, the growth rate of resistant cells after therapy is identical to that of the sensitive cells before treatment,  $d = 0.5b$  in the absence of treatment,  $M = 10^{-6}$ , and treatment begins when the tumour has  $N = 10^7$  cells.

Note that our model assumes the drug is uniformly distributed in the tumour<sup>96</sup>; it is known that drug gradients can speed up the onset of resistance<sup>101</sup>.

Supplementary Video 1 and Extended Data Fig. 4a show that, since the process of resistance acquisition is stochastic, some tumours regrow after an initial

regression, and some do not. If only resistant cells can migrate, regrowth is faster (Extended Data Fig. 4b, c). Extended Data Fig. 4d–g shows regrowth probabilities  $P_{\text{regrowth}}$  for different treatment scenarios not mentioned in the main text, depending on whether the drug is cytostatic ( $b_{\text{treatment}} = 0$ ) or cytotoxic ( $d_{\text{treatment}} = b$ ), and whether  $d = 0$  or  $d > 0$  before treatment. In Extended Data Fig. 4d, cells replicate and die only on the surface, and the core is ‘quiescent’—cells are still alive there but cannot replicate unless outer layers are removed by treatment (Supplementary Videos 6 and 7).  $P_{\text{regrowth}}$  does not depend on the dispersal probability  $M$  at all, and is close to 100% for  $N > 10^8$  cells, a size that is larger than for  $d > 0$  (Extended Data Fig. 4f). It can be shown that  $P_{\text{regrowth}} = 1 - \exp(-\gamma_i N)$ . Extended Data Fig. 4e is for the cytostatic drug ( $b_{\text{treatment}} = d_{\text{treatment}} = 0$ ); this is also equivalent to the cytotoxic drug if the tumour has a necrotic core (cells are dead but still occupy physical volume). In this case,  $P_{\text{regrowth}}$  increases with  $M$  because more resistant cells are on the surface for larger  $M$  (cells can replicate only on the surface in this scenario). Extended Data Fig. 4f, g shows models with cell death present even in the absence of treatment ( $d = 0.9b$ ) but occurring only at the surface, unlike in Fig. 3 where cells also die inside the tumour. Death increases  $P_{\text{regrowth}}$  owing to a larger number of cellular division necessary to obtain the same size, and hence more opportunities to mutate.

**Relaxing the assumptions of the model.** Figure 4 shows that even a small fitness advantage substantially reduces genetic diversity through the process of clonal expansion, see also Supplementary Videos 4 and 5. We now demonstrate that this also applies to modified versions of the model, proving its robustness.

**Exact values of  $M$  and  $s$  has no qualitative effect.** Extended Data Fig. 8b, e shows that the average number of shared genetic alterations is larger in the presence of drivers also in the case of non-zero dispersal ( $M > 0$ ), and its numerical value is almost the same as for  $M = 0$  (Fig. 4). Extended Data Fig. 8c, f shows that as long as  $s > 0$  and regardless of its exact value, driver mutations reduce genetic diversity in the tumour compared to the case  $s = 0$ . Extended Data Fig. 5a–c shows how many driver mutations are expected to be present in a randomly chosen cell from a tumour that is  $T$  years old. Neither dispersal nor the way drivers affect growth (via birth or death rate) has a significant effect on the number of drivers per cell (Extended Data Fig. 5b, c). A small discrepancy visible in Extended Data Fig. 5b is caused by a slightly asymmetric way death and birth is treated in our model, see the Supplementary Information.

**Model B.** Cells replicate with constant rate if there is at least one empty neighbour. In the absence of drivers, genetic alterations are distributed evenly throughout the lesion (Extended Data Fig. 6b) but they often occur independently and the number of frequent genetic alterations is low (Extended Data Fig. 6c). Drivers cause clonal expansion as in model A.

**Model C.** Cells replicate regardless of whether there are empty sites surrounding them or not. When a cell replicates, it pushes away other cells towards the surface (Supplementary Information). Extended Data Fig. 6c, e shows that this again leads to clonal expansion which decreases diversity.

**Model D.** Replication/death occurs only on the surface and the core of the tumour is static. Extended Data Fig. 6d shows that driver mutations cannot spread to the other side of the lesion and conical clonal sectors can be seen even for  $s > 0$ . The number of frequent genetic alterations is the same for  $s = 0$  and  $s = 1\%$ , indicating that genetic heterogeneity is not lowered by clonal expansion. This demonstrates that cell turnover inside the tumour is very important for reducing heterogeneity. To obtain the same (low) heterogeneity as for models a–c, the probability of driver mutations must be much larger in model D (Extended Data Fig. 6f).

**Drivers affecting  $M$ .** We investigated three scenarios in which drivers affect (1) only the dispersal probability  $M \rightarrow (1 + q)M$ , in which  $q > 0$  is the ‘migration fitness advantage’ (no change in  $b$ ,  $d$ ), (2) both  $M$  and  $d$ , that is,  $(d, M) \rightarrow (d(1 - s), (1 + q)M)$  with  $s, q > 0$ , (3) either  $M$  or  $d$ , with probability 1/2. Extended Data Fig. 3c shows that growth is unaffected in cases (1, 3) compared to the neutral case. For (2) the tumour growth rate increases significantly when the tumour is larger than  $N = 1 \times 10^6$  cells. This shows that migration increases the overall fitness advantage, in line with ref. 102, which shows that fixation probability is determined by the product of the exponential growth rate and diffusion constant (motility) of organisms.

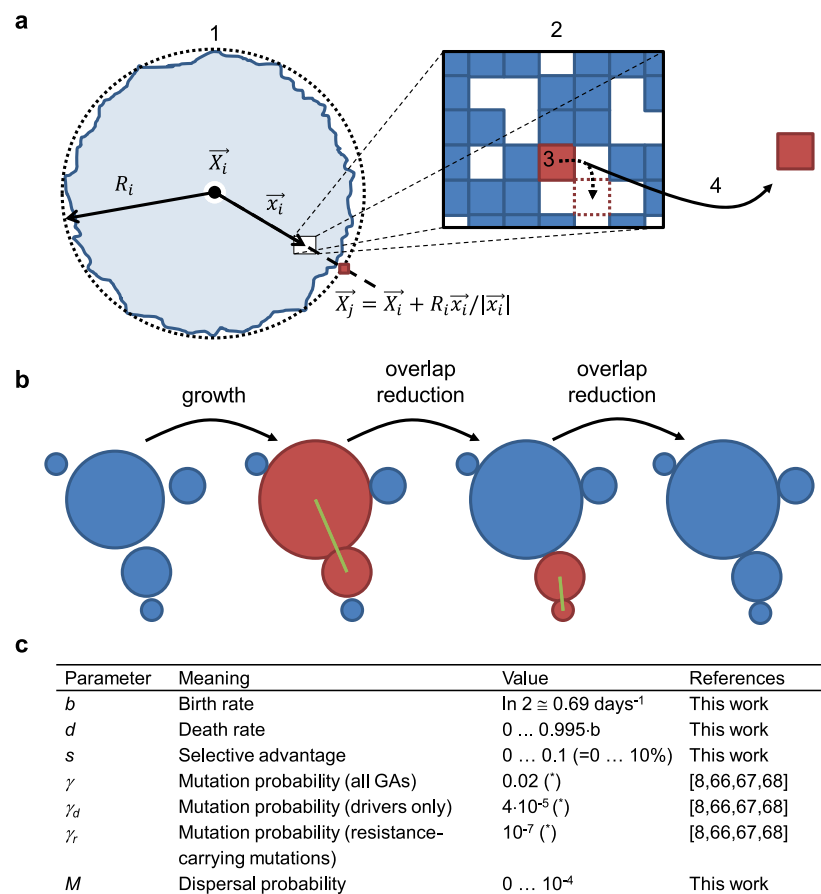
**Six-site (von Neumann) neighbourhood.** We simulated a model in which each cell has only six neighbours (von Neumann neighbourhood) instead of 26 (Moore neighbourhood). Extended Data Fig. 9 compares models A and C for the two neighbourhoods and show that there is only a small quantitative difference in the growth curves for model A (model C is unaffected), but that the shape of the ball of cells deviates more from the spherical one for the six-site neighbourhood, see also section 7 in the Supplementary Information.

- Rodríguez-Brenes, I. A., Komarova, N. L. & Wodarz, D. Tumor growth dynamics: insights into somatic evolutionary processes. *Trends Ecol. Evol.* **28**, 597–604 (2013).



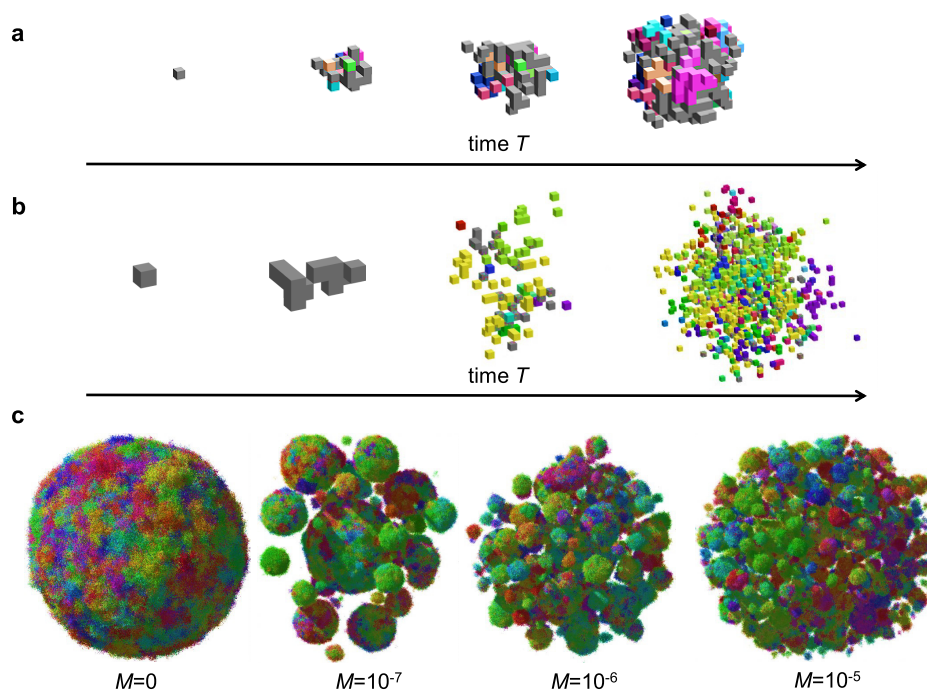
32. Anderson, A. R. A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Math. Med. Biol.* **22**, 163–186 (2005).
33. Komarova, N. L. Spatial stochastic models for cancer initiation and progression. *Bull. Math. Biol.* **68**, 1573–1599 (2006).
34. Lavrentovich, M. O. & Nelson, D. R. Survival probabilities at spherical frontiers. *Theor. Popul. Biol.* **102**, 26–39 (2015).
35. Komarova, N. L. Spatial stochastic models of cancer: fitness, migration, invasion. *Math. Biosci. Eng.* **10**, 761–775 (2013).
36. Manem, V. S. K., Kohandel, M., Komarova, N. L. & Sivaloganathan, S. Spatial invasion dynamics on random and unstructured meshes: implications for heterogeneous tumor populations. *J. Theor. Biol.* **349**, 66–73 (2014).
37. Durrett, R., Foo, J. & Leder, K. Spatial Moran models, II: cancer initiation in spatially structured tissue. *J. Math. Biol.* <http://dx.doi.org/10.1007/s00285-015-0912-1> (2015).
38. Gerlee, P. & Nelander, S. The impact of phenotypic switching on glioblastoma growth and invasion. *PLOS Comput. Biol.* **8**, e1002556 (2012).
39. González-García, I., Solé, R. V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *Proc. Natl Acad. Sci. USA* **99**, 13085–13089 (2002).
40. Sehyo, C. C. *et al.* Model for *in vivo* progression of tumors based on co-evolving cell population and vasculature. *Sci. Rep.* **1**, 31 (2011).
41. Torquato, S. Toward an Ising model of cancer and beyond. *Phys. Biol.* **8**, 015017 (2011).
42. Reiter, J. G., Bozic, I., Allen, B., Chatterjee, K. & Nowak, M. A. The effect of one additional driver mutation on tumor progression. *Evol. Appl.* **6**, 34–45 (2013).
43. Kansal, A. R., Torquato, S., Harsh, G. R., Chiocca, E. A. & Deisboeck, T. S. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *J. Theor. Biol.* **203**, 367–382 (2000).
44. Kansal, A. R., Torquato, S., Chiocca, E. A. & Deisboeck, T. S. Emergence of a subpopulation in a computational model of tumor growth. *J. Theor. Biol.* **207**, 431–441 (2000).
45. Antal, T., Krapivsky, P. L. & Nowak, M. A. Spatial evolution of tumors with successive driver mutations. *Phys. Rev. E* **92**, 022705 (2015).
46. Enderling, H., Hlatky, L. & Hahnfeldt, P. Migration rules: tumours are conglomerates of self-metastases. *Br. J. Cancer* **100**, 1917–1925 (2009).
47. Sottoriva, A. *et al.* Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer Res.* **70**, 46–56 (2010).
48. Schaller, G. & Meyer-Hermann, M. Multicellular tumor spheroid in an off-lattice voronoi-delaunay cell model. *Phys. Rev. E* **71**, 051910 (2005).
49. Radszweit, M., Block, M., Hengstler, J. G., Schöll, E. & Drasdo, D. Comparing the growth kinetics of cell populations in two and three dimensions. *Phys. Rev. E* **79**, 051907 (2009).
50. Moglia, B., Guisoni, N. & Albano, E. V. Interfacial properties in a discrete model for tumor growth. *Phys. Rev. E* **87**, 032713 (2013).
51. Foo, J., Leder, K. & Ryser, M. Multifocality and recurrence risk: a quantitative model of field cancerization. *J. Theor. Biol.* **355**, 170–184 (2014).
52. Poleszczuk, J., Hahnfeldt, P. & Enderling, H. Evolution and phenotypic selection of cancer stem cells. *PLOS Comput. Biol.* **11**, e1004025 (2015).
53. Durrett, R., Schmidt, D. & Schweinsberg, J. A waiting time problem arising from the study of multi-stage carcinogenesis. *Ann. Appl. Probab.* **19**, 676–718 (2009).
54. Spencer, S. L., Berryman, M. J., García, J. A. & Abbott, D. An ordinary differential equation model for the multistep transformation to cancer. *J. Theor. Biol.* **231**, 515–524 (2004).
55. Kim, Y. & Othmer, H. G. A hybrid model of tumor–stromal interactions in breast cancer. *Bull. Math. Biol.* **75**, 1304–1350 (2013).
56. Ramis-Conde, I., Chaplain, M. A. J., Anderson, A. R. A. & Drasdo, D. Multi-scale modelling of cancer cell intravasation: the role of cadherins in metastasis. *Phys. Biol.* **6**, 016008 (2009).
57. Swanson, K. R. *et al.* Quantifying the role of angiogenesis in malignant progression of gliomas: *in silico* modeling integrates imaging and histology. *Cancer Res.* **71**, 7366–7375 (2011).
58. Taloni, A. *et al.* Mechanical properties of growing melanocytic nevi and the progression to melanoma. *PLoS ONE* **9**, e94229 (2014).
59. Drasdo, D. & Höhme, S. A single-cell-based model of tumor growth in vitro: monolayers and spheroids. *Phys. Biol.* **2**, 133 (2005).
60. Drasdo, D., Hoehme, S. & Block, M. On the role of physics in the growth and pattern formation of multi-cellular systems: what can we learn from individual-cell based models? *J. Stat. Phys.* **128**, 287–345 (2007).
61. Jiang, Y., Pjesivac-Grbovic, J., Cantrell, C. & Freyer, J. P. A multiscale model for avascular tumor growth. *Biophys. J.* **89**, 3884–3894 (2005).
62. Eden, M. in *A Two-Dimensional Growth Process* (eds Family, F. & Vicsek, T.) 265–283 (World Scientific, 1961).
63. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates., 1997).
64. Kreft, J. U., Booth, G. & Wimpenny, J. W. T. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* **144**, 3275–3287 (1998).
65. Lardon, L. A. *et al.* iDynoMiCS: Next-generation individual-based modelling of biofilms. *Environ. Microbiol.* **13**, 2416–2434 (2011).
66. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105**, 4283–4288 (2008).
67. Wang, T. L. *et al.* Prevalence of somatic alterations in the colorectal cancer cell genome. *Proc. Natl Acad. Sci. USA* **99**, 3076–3080 (2002).
68. Diaz, L. A. Jr *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
69. Honda, H., Morita, T. & Tanabe, A. Establishment of epidermal cell columns in mammalian skin: computer simulation. *J. Theor. Biol.* **81**, 745–759 (1979).
70. Ali, A., Somfai, E. & Grosskinsky, S. Reproduction-time statistics and segregation patterns in growing populations. *Phys. Rev. E* **85**, 021923 (2012).
71. Korolev, K. S., Xavier, J. B., Nelson, D. R. & Foster, K. R. Data from: a quantitative test of population genetics using spatio-genetic patterns in bacterial colonies. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.3147q> (2011).
72. Gong, P., Wang, Y., Liu, G., Zhang, J. & Wang, Z. New insight into Ki67 expression at the invasive front in breast cancer. *PLoS ONE* **8**, e54912 (2013).
73. Ellison, T. A. *et al.* A single institution's 26-year experience with nonfunctional pancreatic neuroendocrine tumors: a validation of current staging systems and a new prognostic nomogram. *Ann. Surg.* **259**, 204–212 (2014).
74. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105**, 4283–4288 (2008).
75. Lindstrom, L. S. *et al.* Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J. Clin. Oncol.* **30**, 2601–2608 (2012).
76. Voss, M. H. *et al.* Tumor genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR inhibitor therapy. *Clin. Cancer Res.* **20**, 1955–1964 (2014).
77. Wiser, M. J., Ribbeck, N. & Lenski, R. E. Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367 (2013).
78. White, T. C. Increased mRNA Levels of *ERG16*, *CDR*, and *MDR1* correlate with increases in azole resistance in *Candida albicans* isolates from a patient infected with human immunodeficiency virus. *Antimicrob. Agents Chemother.* **41**, 1482–1487 (1997).
79. McGranahan, N. *et al.* Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci. Transl. Med.* **7**, 283 (2015).
80. Ranft, J. *et al.* Fluidization of tissues by cell division and apoptosis. *Proc. Natl Acad. Sci. USA* **107**, 20863–20868 (2010).
81. LeBleu, V. S. *et al.* PGC-1 $\alpha$  mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nature Cell Biol.* **16**, 992–1003 (2014).
82. Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).
83. Sciumè, G. *et al.* A multiphase model for three-dimensional tumor growth. *New J. Phys.* **15**, 015005 (2013).
84. Charras, G. & Sahai, E. Physical influences of the extracellular environment on cell migration. *Nature Rev. Mol. Cell Biol.* **15**, 813–824 (2014).
85. Jiao, Y. & Torquato, S. Diversity of dynamics and morphologies of invasive solid tumors. *AIP Advances* **2**, 011003 (2012).
86. Galle, J., Loeffler, M. & Drasdo, D. Modeling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations *in vitro*. *Biophys. J.* **88**, 62–75 (2005).
87. Chen, E. J., Novakofski, J., Jenkins, W. K. & O'Brien, W. D. Jr. Young's modulus measurements of soft tissues with application to elasticity imaging. Ultrasonics, ferroelectrics, and frequency control. *IEEE Transactions* **43**, 191–194 (1996).
88. Samani, A., Bishop, J., Luginbuhl, C. & Plewes, D. B. Measuring the elastic modulus of *ex vivo* small tissue samples. *Phys. Med. Biol.* **48**, 2183 (2003).
89. Weinberg, R. A. *The Biology of Cancer* (Garland Science, 2007).
90. Lekka, M. *et al.* Elasticity of normal and cancerous human bladder cells studied by scanning force microscopy. *Eur. Biophys. J.* **28**, 312–316 (1999).
91. Gonzalez-Rodriguez, D., Guevorkian, K., Douezan, S. & Brochard-Wyart, F. Soft matter models of developing tissues and tumors. *Science* **338**, 910–917 (2012).
92. Stirbat, T. V. *et al.* Fine tuning of tissues' viscosity and surface tension through contractility suggests a new role for  $\alpha$ -catenin. *PLoS ONE* **8**, e52554 (2013).
93. Drasdo, D. Buckling instabilities of one-layered growing tissues. *Phys. Rev. Lett.* **84**, 4244–4247 (2000).
94. Basan, M., Joanny, J.-F., Prost, J. & Risler, T. Undulation instability of epithelial tissues. *Phys. Rev. Lett.* **106**, 158101 (2011).
95. Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* **2**, e00747 (2013).
96. Stylianopoulos, T. & Jain, R. K. Combining two strategies to improve perfusion and drug delivery in solid tumors. *Proc. Natl Acad. Sci. USA* **110**, 18632–18637 (2013).
97. Foo, J. & Michor, F. Evolution of acquired resistance to anti-cancer therapy. *J. Theor. Biol.* **355**, 10–20 (2014).
98. Goldie, J. H. & Coldman, A. J. The genetic origin of drug resistance in neoplasms: implications for systemic therapy. *Cancer Res.* **44**, 3643–3653 (1984).
99. Coldman, A. J. & Goldie, J. H. A stochastic model for the origin and treatment of tumors containing drug-resistant cells. *Bull. Math. Biol.* **48**, 279–292 (1986).
100. Coldman, A. J. & Goldie, J. H. A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Math. Biosci.* **65**, 291–307 (1983).
101. Gremlich, P., Wacław, B. & Allen, R. J. Mutational pathway determines whether drug gradients accelerate evolution of drug-resistant cells. *Phys. Rev. Lett.* **109**, 088101 (2012).
102. Korolev, K. S. *et al.* Selective sweeps in growing microbial colonies. *Phys. Biol.* **9**, 026008 (2012).





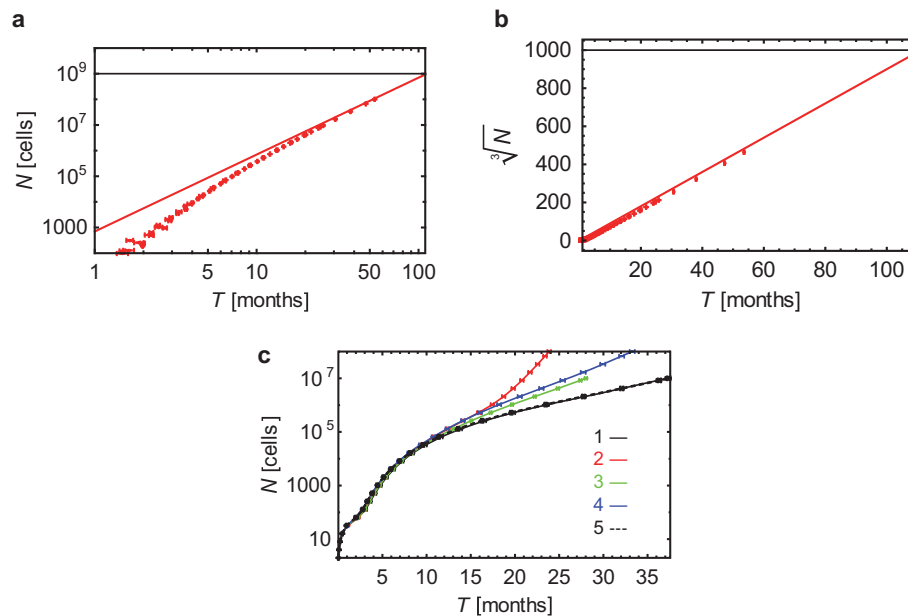
**Extended Data Figure 1 | Details of the model.** **a**, A sketch showing how dispersal is implemented: (1) A ball of cells of radius  $R_i$ , in which the centre is at  $X_i$ , is composed of tumour cells and normal cells (blue and empty squares in the zoomed-in rectangle (2)). A cell at position  $x_i$  with respect to the centre of the ball attempts to replicate (3). If replication is successful, the cell migrates with probability  $M$  and creates a new microlesion (4). The position  $X_j$  of this new ball of cells is determined as the endpoint of the vector that starts at  $X_i$  and has direction  $x_i$  and length  $R_i$ . **b**, Overlap reduction between the balls of cells. When a growing ball begins to overlap with another ball (red), they

are both moved apart along the line connecting their centres of mass (green line) by as much as necessary to reduce the overlap to zero. The process is repeated for all overlapping balls as many times as needed until there is no overlap. **c**, Summary of all parameters used in the model. If, for a given parameter, many different values have been used in different plots, a range of values used is shown. Birth and death rates can also depend on the number of driver mutations, see Methods. Asterisk, parameter estimated from other quantities available in the literature.



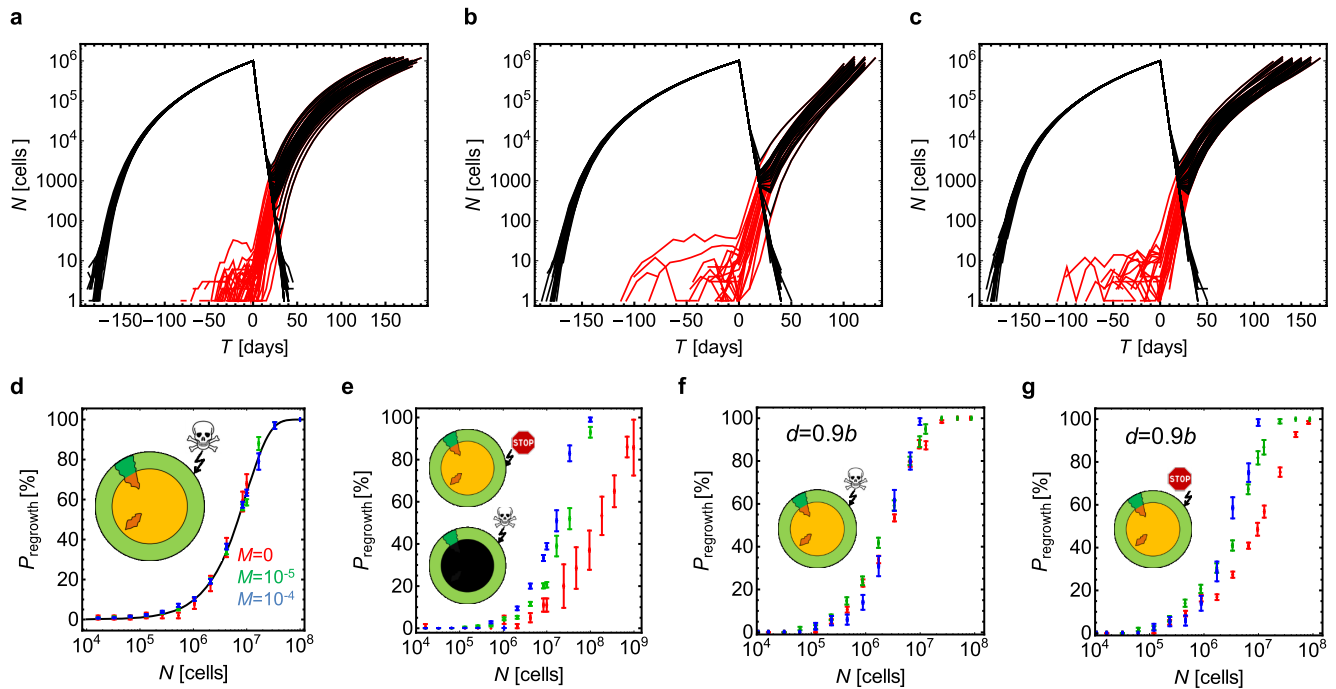
**Extended Data Figure 2 | Simulation snapshots.** **a, b,** A few snapshots of tumour growth for no dispersal, and  $d=0$  (**a**) and  $d=0.9b$  (**b**). To visualize clonal sectors, cells have been colour-coded by making the colour a heritable trait and changing each of its RGB components by a small random fraction whenever a cell mutates. The initial cell is grey. Empty space (white)

are non-cancer cells mixed with extracellular matrix. Note that images are not to scale. **c,** Tumour shapes for  $N=1 \times 10^7$ ,  $d=0.9b$ , and different dispersal probability  $M$ . Images not to scale; the tumour for  $M=1 \times 10^{-5}$  is larger than the one for  $M=0$ .



**Extended Data Figure 3 | Tumour size as a function of time.** **a**, Growth of a tumour without dispersal ( $M = 0$ ), for  $d = 0.8b$ . For large times ( $T$ ), the number of cells grows approximately as  $\text{const} \times T^3$ . The tumour reaches size  $N = 1 \times 10^9$  cells (horizontal line) after about 100 months (8 years) of growth. **b**, The same data are plotted in the linear scale, with  $N$  replaced by 'linear extension'  $N^{1/3}$ . **c**, Tumour size versus time when drivers affect the dispersal probability. In all cases,  $d = 0.9b$ , and (1, black) drivers increase the dispersal rate tenfold ( $q = 9$ ) but have no effect on the net growth rate; (2, red)

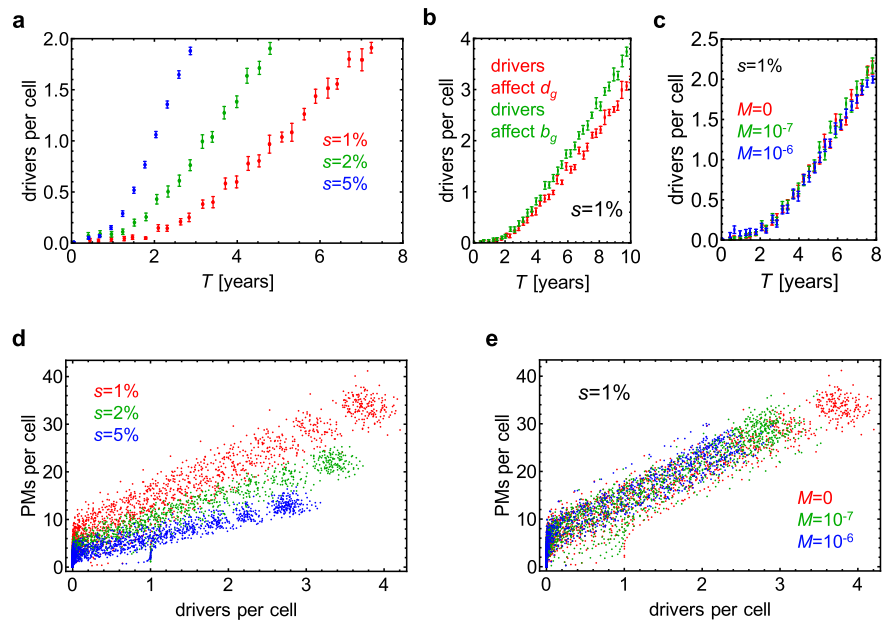
drivers increase both the net growth rate ( $s = 10\%$ ) and  $M$ ; (3, green) drivers either (with probability 1/2) increase  $M$  tenfold ( $q = 9$ ) or increase the net growth rate by  $s = 10\%$ ; (4, blue) drivers increase only the net growth rate by  $s = 10\%$ ; and (5, black dashed line) neutral case with  $M = 1 \times 10^{-7}$ , which is indistinguishable from (1). In all cases (1–3) the initial dispersal probability  $M = 1 \times 10^{-7}$ . Points represent average value over 40–100 simulations per data point, error bars are s.e.m.



**Extended Data Figure 4 | Simulation of targeted therapy.** **a–c**, The total number of cells in the tumour (black) and the number of resistant cells (red) versus time, during growth ( $T < 0$ ) and treatment ( $T > 0$ ), for  $\sim 100$  independent simulations, for  $d = 0.5b$  for  $T < 0$ . Therapy begins when  $N = 1 \times 10^6$  cells. After treatment, many tumours die out ( $N$  decreases to zero) but those with resistant cells will regrow sooner or later. **a**,  $M = 0$  for all cells at all times. **b**,  $M = 0$  for all cells for  $T < 0$  and  $M = 10^{-4}$  for resistant cells for  $T > 0$ . **c**,  $M = 0$  for non-resistant and  $M = 10^{-5}$  for resistant cells at all times. In all three cases,  $P_{\text{regrowth}}$  is very similar:  $36 \pm 5\%$  (mean  $\pm$  s.e.m.) (**a**),  $25 \pm 4\%$  (**b**), and  $27 \pm 6\%$  for (**c**). **d–g**, Regrowth probability for four treatment scenarios not discussed in the main text. Data points correspond to three dispersal probabilities:  $M = 0$  (red),  $M = 1 \times 10^{-5}$  (green), and  $M = 1 \times 10^{-4}$  (blue).

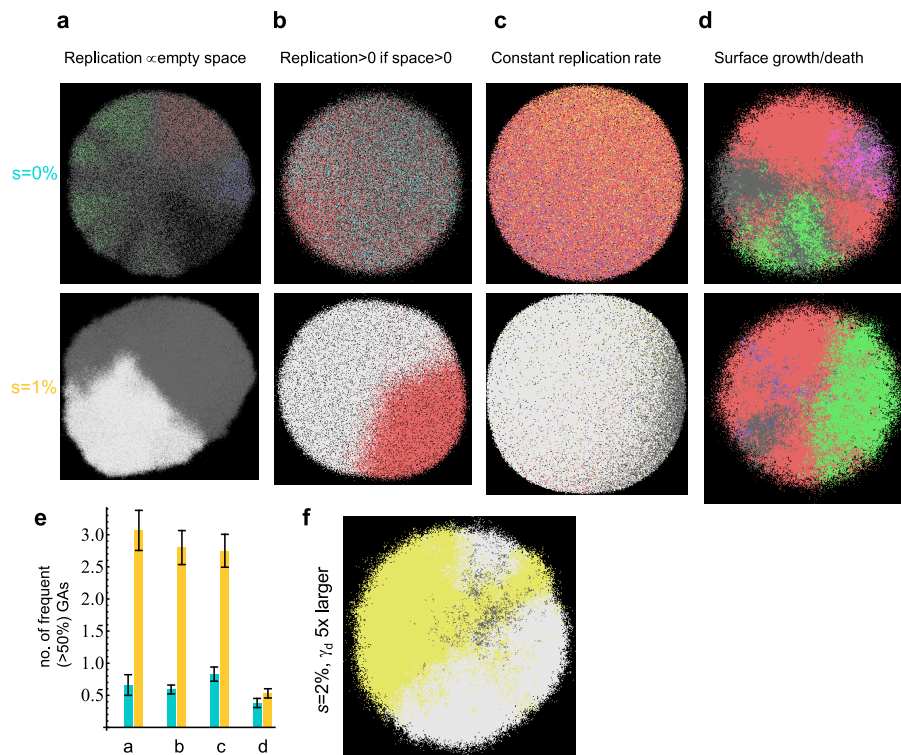
The probability is plotted as a function of tumour size  $N$  just before the therapy commences. **d**, Before treatment, cells replicate only on the surface. Cells in the core are quiescent and do not replicate. Therapy kills cells on the surface and cells in the core resume proliferation when liberated by treatment. **e**, As in **d**, but drug is cytostatic and does not kill cells but inhibits their growth. The results for  $P_{\text{regrowth}}$  are identical if the drug is cytotoxic and the tumour has a necrotic core (cells die inside the tumour and cannot replicate even if the surface is removed). **f**, Before treatment, cells replicate and die on the surface. The core is quiescent. Therapy kills cells on the surface (cytotoxic drug). **g**, As in **f**, but therapy only inhibits growth (cytostatic drug). In all cases (**d–g**) error bars represent s.e.m. from 8–1,000 simulations per point.





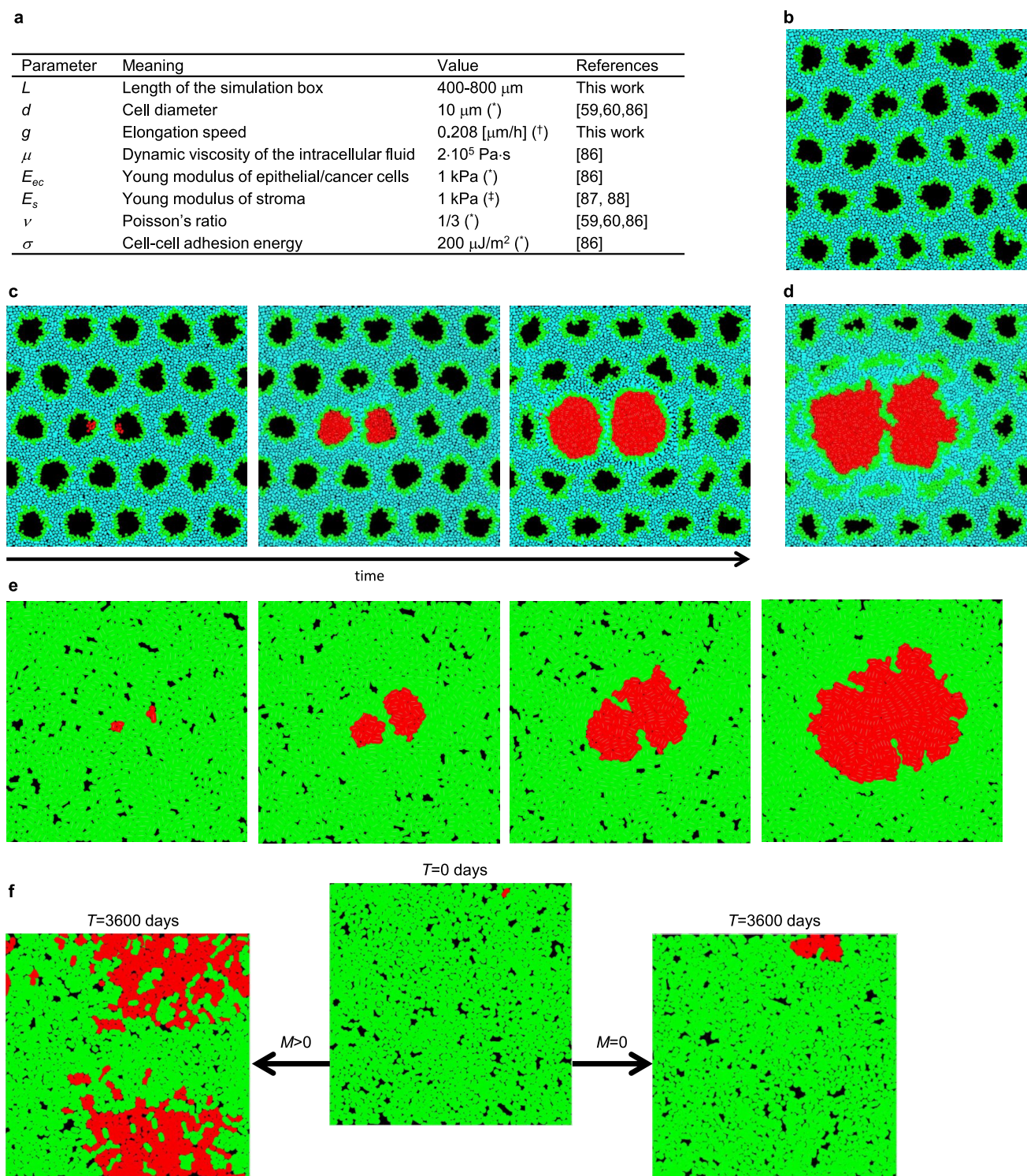
**Extended Data Figure 5 | Accumulation of driver and passenger genetic alterations.** **a–c,** The number of drivers per cell in the primary tumour plotted as a function of time (10–100 simulations per point, error bars denote s.e.m.). **a,**  $M = 0$  and three different driver selective advantages. For  $s = 1\%$ , cells accumulate on average one driver mutation within 5 years. The time can be significantly lower for very strong drivers ( $s > 1\%$ ). **b,** The rate at which drivers accumulate depends mainly on their selective advantage and not on whether

they affect death or birth rate. **c,** Dispersal does not affect the rate of driver accumulation. **d, e,** The number of passenger mutations (PMs) per cell versus the number of driver mutations per cell. More passenger mutations are present for smaller driver selective advantage (**d**), and this is independent of the dispersal probability  $M$  (**e**) in the regime of small  $M$ . Data points correspond to independent simulations.



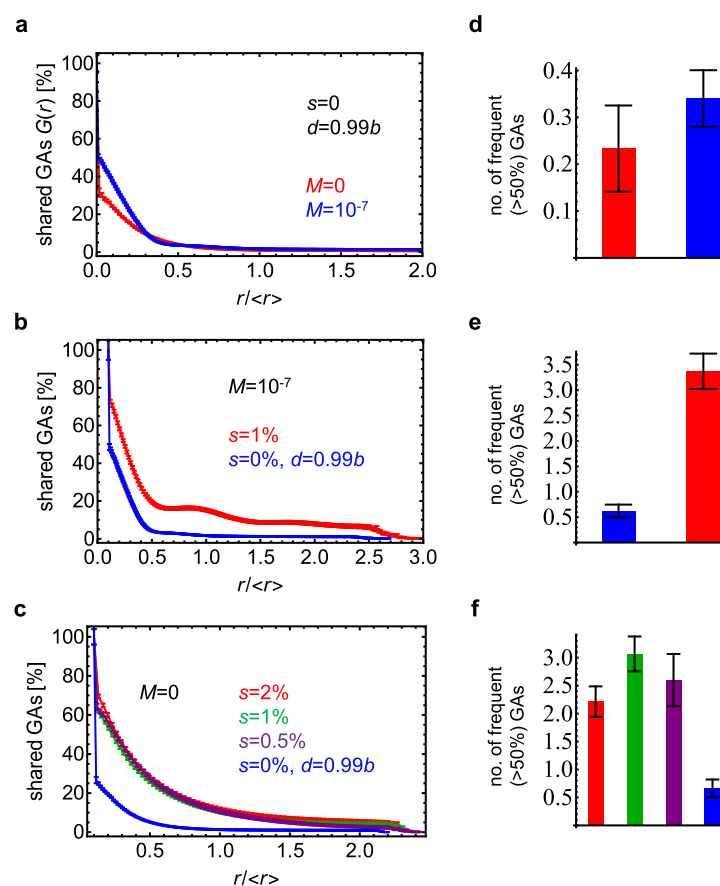
**Extended Data Figure 6 | Genetic diversity in a single lesion for different models.** **a–d**, Representative simulation snapshots, with genetic alterations colour-coded as in Fig. 4. Top:  $s = 0$ , bottom:  $s = 1\%$ . **a**, Model A from the main text in which cells replicate with rates proportional to the number of empty nearby sites. **b**, Model B, the replication rate is constant and non-zero if there is at least one empty site nearby, and zero otherwise. **c**, Model C, cells replicate at a constant rate and push away other cells to make space for their progeny. **d**, Model D, cells replicate/die only on the surface, the interior of the tumour

(‘necrotic core’) is static. In all cases,  $N = 1 \times 10^7$ ,  $d = 0.99b$ . **e**, Number of genetic alterations present in at least 50% of cells for identical parameters as in **a–d**. In all cases except surface growth (**d**), drivers increase genetic homogeneity, as measured by the number of most frequent genetic alterations. Results averaged over 50–100 simulations, error bars denote s.e.m. **f**, Model D, with  $\gamma_d = 2 \times 10^{-4}$  instead of  $4 \times 10^{-5}$ , that is, drivers occur five times more often. In this case, driver mutations arise earlier than in **d**, and the tumour becomes more homogeneous.



**Extended Data Figure 7 | The off-lattice model.** **a**, Summary of all parameters used in the model. Asterisk, typical value, varies between different types of tissues; dagger symbol, equivalent to 24 h minimal doubling time; double dagger symbol, based on the assumption that macroscopic elastic properties of tissues such as liver, pancreases or mammary glands are primarily determined by the elastic properties of stroma. **b**, Simulation snapshot of a normal tissue before the invasion of cancer cells. **c**, Two balls of cancer cells in

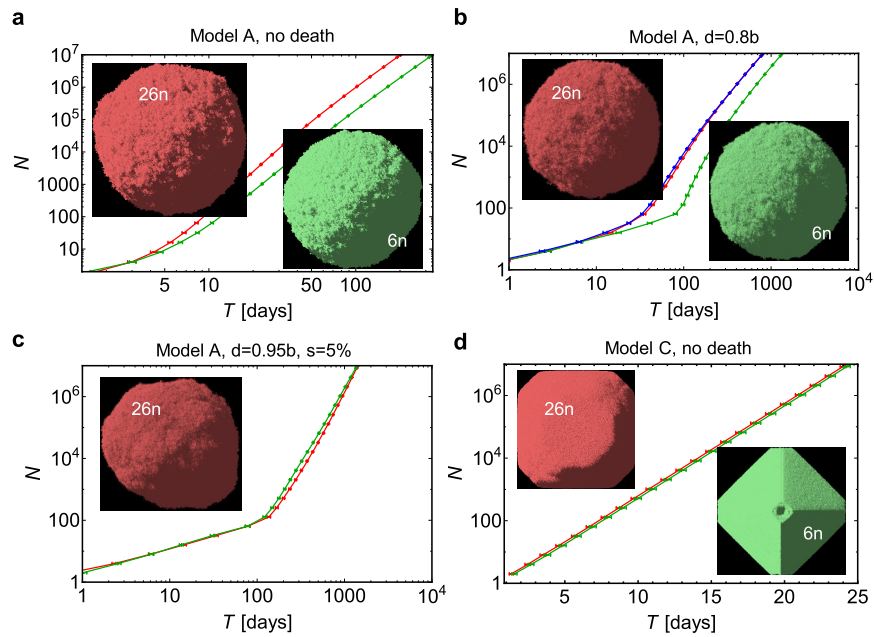
two nearby ducts repel each other as they grow as a consequence of mechanical forces exerted on each other. **d**, The balls coalesce if growth is able to break the separating extracellular matrix. **e**, If the balls are not encapsulated, they quickly merge. **f**, Isolated balls of cells are not required to speed up growth; migration (left) can cause the tumour to expand much faster even if individual microlesions merge together, as opposed to the case with no migration (right).



**Extended Data Figure 8 | Genetic diversity quantified.** **a**, Tumours are much more genetically heterogeneous in the absence of driver mutations ( $s = 0$ ) (see Fig. 4). The plot shows the fraction  $G(r)$  of genetic alterations (GAs) shared between the cells as function of their separation (distance  $r$ ) in the tumour. The fraction quickly decreases with increasing  $r$ . The distance in the figure is normalized by the average distance  $\langle r \rangle$  between any two cells in the tumour. For a spherical tumour,  $\langle r \rangle$  is approximately equal to half of the tumour diameter. **b**, Fraction of shared genetic alterations for  $s = 1\%$  and  $s = 0\%$ ,

$N = 1 \times 10^7$ , and  $M = 1 \times 10^{-7}$ . In the presence of drivers,  $G(r)$  decays slower, indicating more homogeneous tumours. **c**, The exact value of the selective advantage of driver mutations is not important (all curves  $G(r)$  look the same, except for  $s = 0$ ) as long as  $s > 0$ . **d–f**, Number of genetic alterations present in at least 50% of cells for identical parameters as in **a–c**, correspondingly. Drivers substantially increase the level of genetic homogeneity. In all panels the results have been averaged over 30–100 simulations, with error bars as s.e.m.





**Extended Data Figure 9 | Growth curves for the 26-nearest neighbours (26n, red curves) and the 6-nearest neighbours (6n, green curves) models.** **a**, Model A (as in the main text), no death. The tumour grows about twice as slow in the 6n model. Pictures show tumour snapshots for both models; there is no visible difference in the shape. **b**, Model A, death  $d = 0.8b$ . The additional blue curve is for the 6n model, with modified replication probability to account for missing neighbours as explained in the Supplementary Information. **c**, Model A, with death  $d = 0.95b$ , and drivers  $s = 5\%$ . There is

very little difference in the growth curves between the 6n and 26n models. A small asymmetry in the shape is caused by faster-growing cells with driver mutations. **d**, Model C (exponential growth). Growth is the same in both 6n and 26n models, but the shape is more aspheric for the 6n model. This is probably caused by shifting cells along the coordinate axes and not along the shortest path to the surface when making space for new cells. All plots show the mean (average over 50–100 simulations) and s.e.m.

**Extended Data Table 1 | Experimental results for the percentage of proliferating cells in the centre versus the edge of solid tumours**

Case	Tumor type	Edge			Center			Ratio center:edge	p-value
		Images	KI %	Total no. of cells	Images	KI %	Total no. of cells		
1	Chromophobe RCC	8	3.46	1013	3	1.11	2561	0.32	0.05
2	Chromophobe RCC	5	3.07	508	3	1.05	938	0.34	0.28
3	Chromophobe RCC	5	2.63	524	3	0.44	697	0.17	0.03
4	Chromophobe RCC	7	1.58	581	2	0.53	958	0.34	0.17
5	HCC	7	17.14	892	2	9.74	1637	0.57	0.05
6	HCC	7	51.71	1079	4	32.84	2562	0.64	0.03
7	HCC	6	47.37	435	3	19.97	1397	0.42	0.09
8	HCC	7	19.02	895	4	13.78	1191	0.72	0.35
9	HCC	6	15.09	1074	3	11.98	1094	0.79	0.33
10	HCC	9	29.84	1305	2	20.87	2457	0.70	0.22
Summary									
1-4	Chromophobe RCC	25	2.69	—	11	0.81	—	0.30	0.00002
5-10	HCC	42	30.0	—	18	19.1	—	0.64	0.007

A representative section of each tumour was labelled for the proliferation marker Ki67 (KI), and images of the tumour at the leading edge and the centre were acquired as described (Methods). Proliferation is markedly increased at the leading edge, and this is statistically significant ('Summary', Kolmogorov–Smirnov two-sample test,  $P < 0.05$ ). The average ratio of the number of proliferating cells in the centre/at the edge is 0.50 (range 0.17–0.79). HCC, hepatocellular carcinoma; RCC, renal cell carcinoma

# Allosteric receptor activation by the plant peptide hormone phytosulfokine

Jizong Wang<sup>1\*</sup>, Hongju Li<sup>2\*</sup>, Zhifu Han<sup>1</sup>, Heqiao Zhang<sup>1</sup>, Tong Wang<sup>2</sup>, Guangzhong Lin<sup>1</sup>, Junbiao Chang<sup>3</sup>, Weicai Yang<sup>2</sup> & Jijie Chai<sup>1</sup>

Phytosulfokine (PSK) is a disulfated pentapeptide that has a ubiquitous role in plant growth and development<sup>1,2</sup>. PSK is perceived by its receptor PSKR<sup>3,4</sup>, a leucine-rich repeat receptor kinase (LRR-RK). The mechanisms underlying the recognition of PSK, the activation of PSKR and the identity of the components downstream of the initial binding remain elusive. Here we report the crystal structures of the extracellular LRR domain of PSKR in free, PSK- and co-receptor-bound forms. The structures reveal that PSK interacts mainly with a  $\beta$ -strand from the island domain of PSKR, forming an anti- $\beta$ -sheet. The two sulfate moieties of PSK interact directly with PSKR, sensitizing PSKR recognition of PSK. Supported by biochemical, structural and genetic evidence, PSK binding enhances PSKR heterodimerization with the somatic embryogenesis receptor-like kinases (SERKs). However, PSK is not directly involved in PSKR–SERK interaction but stabilizes PSKR island domain for recruitment of a SERK. Our data reveal the structural basis for PSKR recognition of PSK and allosteric activation of PSKR by PSK, opening up new avenues for the design of PSKR-specific small molecules.

Peptide signalling has critical roles in regulating plant physiology<sup>1</sup>. Phytosulfokine (PSK)<sup>5</sup> is a secreted disulfated pentapeptide (Tyr(SO<sub>3</sub>H)-Ile-Tyr(SO<sub>3</sub>H)-Thr-Gln) that has ubiquitous roles in plant growth and development<sup>2</sup>. PSK matures through proteolytic cleavage of its precursor proteins<sup>6</sup> with post-translational sulfation<sup>7</sup> for its full activity<sup>8,9</sup>. PSK receptor was first identified in *Daucus carota* (carrot)<sup>3</sup> and the corresponding gene *DcPSKR* is conserved among plants including *Arabidopsis* that encodes two *PSKR* orthologues, *PSKR1* (ref. 4) and *PSKR2* (ref. 10), but PSK perception largely relies on *PSKR1* (ref. 10).

*DcPSKR* and *PSKR1/2* (PSKRs) belong to the large family of leucine-rich repeat receptor kinases (LRR-RKs) with an extracellular LRR domain and a cytoplasmic kinase domain (KD)<sup>11</sup>. The extracellular domains of the three LRR-RKs contain 21 LRRs with an island domain (ID) required for PSK perception<sup>3,10</sup>. PSK binding induces signalling mediated by Ca<sup>2+</sup>/CaM binding and the kinase activity of *PSKR1* (ref. 12), suggesting that ligand binding activates the *PSKR1*<sup>KD</sup>, as observed in the well-studied RKs such as flagellin insensitive 2 (FLS2) and brassinosteroid insensitive 1 (BRI1)<sup>13</sup>. Signalling mediated by the latter two receptor kinases requires ligand-enhanced heterodimerization with the LRR-RK BAK1 (ref. 14), a member of somatic embryogenesis receptor-like kinases (SERKs) that generally act as a co-receptor with other LRR-RKs<sup>15</sup>.

We first solved the crystal structures of the PSK–*PSKR1*<sup>LRR</sup> (Fig. 1a and Extended Data Table 1) and PSK–*DcPSKR*<sup>LRR</sup> (Extended Data Fig. 1a and Extended Data Table 1) complexes. PSK adopts a  $\beta$ -strand conformation, forming an anti-parallel  $\beta$ -sheet with the *PSKR1*<sup>ID</sup> (Fig. 1a). Besides the hydrogen bonds within the  $\beta$ -sheet (Fig. 1b), *PSKR1*<sup>Ser370</sup>, *PSKR1*<sup>Ser372</sup>, *PSKR1*<sup>Thr398</sup> and *PSKR1*<sup>Asp445</sup> from the inner side of the helical structure also form hydrogen bonds with the

main chain of PSK (Fig. 1c). Additionally, *PSKR1*<sup>Arg300</sup> and *PSKR1*<sup>Asn346</sup> form hydrogen bonds with the free carboxyl group of PSK<sup>Gln5</sup> (Fig. 1c), whereas *PSKR1*<sup>Phe506</sup> tightly packs against PSK<sup>Gln5</sup> and PSK<sup>Tyr3</sup>. The two sulfate moieties contribute to PSK–*PSKR1*<sup>LRR</sup> interactions via both hydrogen bonds involving *PSKR1*<sup>Lys508</sup> and *PSKR1*<sup>Asn424</sup> and van der Waals packing involving *PSKR1*<sup>Leu399</sup>, *PSKR1*<sup>Trp448</sup> and *PSKR1*<sup>Lys508</sup> (Fig. 1b, c). The PSK-interacting residues of *PSKR1* are highly conserved in *DcPSKR* (Extended Data Fig. 1b, c) and *PSKR2* (Extended Data Fig. 1d), suggesting that the three PSKRs are conserved in PSK recognition. Indeed, the structure of PSK–*DcPSKR*<sup>LRR</sup> is almost identical to that of PSK–*PSKR1*<sup>LRR</sup> (Fig. 1d) with a r.m.s.d. (root mean square deviation) of 1.45 Å. Further supporting the sulfate group-mediated PSK–*DcPSKR*<sup>LRR</sup> interactions, microscale thermophoresis (MST) showed that PSK displayed a higher binding affinity with *DcPSKR*<sup>LRR</sup> than the desulfated PSK (dPSK) (Extended Data Fig. 2a), agreeing with the observation that dPSK promotes root elongation of *Arabidopsis* plants but with a lower activity than PSK<sup>8</sup>. Previous studies using microsomal fractions derived from cells showed that PSK–*PSKR* interaction displayed a dissociation constant of 4.2 nM in carrot<sup>3</sup> and 7.7 nM in *Arabidopsis*<sup>4</sup>, approximately 200–370 times stronger than the affinity measured between *DcPSKR*<sup>LRR</sup> and PSK by MST. The precise reason for the affinity difference between cell-based and *in vitro* quantification assays is unclear, but it is possible that interactions between transmembrane or cytoplasmic domains within the cellular context provide an environment more favourable for PSK interaction with its receptor. Assays using MST also confirmed the important role of the critical *DcPSKR*<sup>LRR</sup> residues (Extended Data Fig. 1b, c) in PSK recognition, as their mutations compromised PSK–*DcPSKR*<sup>LRR</sup> association, albeit to varying degrees (Extended Data Fig. 2b).

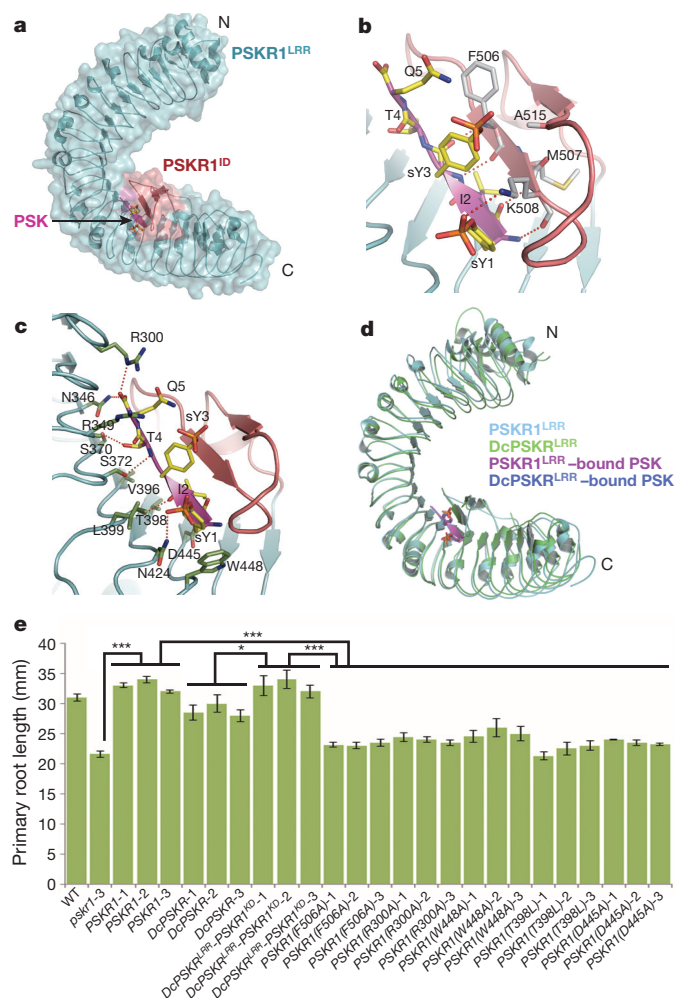
As observed previously<sup>8</sup>, the *pskr1-3 Arabidopsis* mutants displayed a shortened root phenotype (Fig. 1e). The phenotype was fully complemented by wild-type (WT) *PSKR1* and the chimaeric *PSKR1* carrying *DcPSKR*<sup>LRR</sup> and the transmembrane domain and KD of *PSKR1*, and almost fully complemented by *DcPSKR* (Fig. 1e), but not by the *PSKR1* constructs carrying mutations of the residues critical for PSK–*PSKR1* interaction (Fig. 1b, c, e). Furthermore, plants carrying the single *PSKR1* mutants were less responsive to PSK than the WT plants (Extended Data Fig. 2c).

PSK binding induced no oligomerization of *PSKR1*<sup>LRR</sup> or *DcPSKR*<sup>LRR</sup> (Extended Data Fig. 3), suggesting that a co-receptor is required for their activation based on the dimerization model<sup>14</sup>. *PSKR1/2* and *DcPSKR* belong to the same family of LRR-RKs as BRI1 (ref. 11) that utilizes a SERK member as its co-receptor<sup>13</sup>. Moreover, PSK promotes somatic embryogenesis<sup>16</sup>, a marker of which is *DcSERK*<sup>17</sup>. These data prompted us to hypothesize that a SERK member functions as a co-receptor with PSKRs. Indeed, gel filtration showed that PSK induced the formation of a complex between *PSKR1*<sup>LRR</sup> and *SERK1/2/BAK1*<sup>LRR</sup> (Fig. 2a and Extended Data Fig. 4a, b). *PSKR1*<sup>LRR</sup>–*SERK1*<sup>LRR</sup> (Fig. 2a)

<sup>1</sup>Ministry of Education Key Laboratory of Protein Science, Center for Structural Biology, School of Life Sciences, Tsinghua-Peking Joint Center for Life Sciences, Tsinghua University, Beijing 100084, China.

<sup>2</sup>State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. <sup>3</sup>School of Chemistry and Molecular Engineering, Zhengzhou University, Zhengzhou 450001, China.

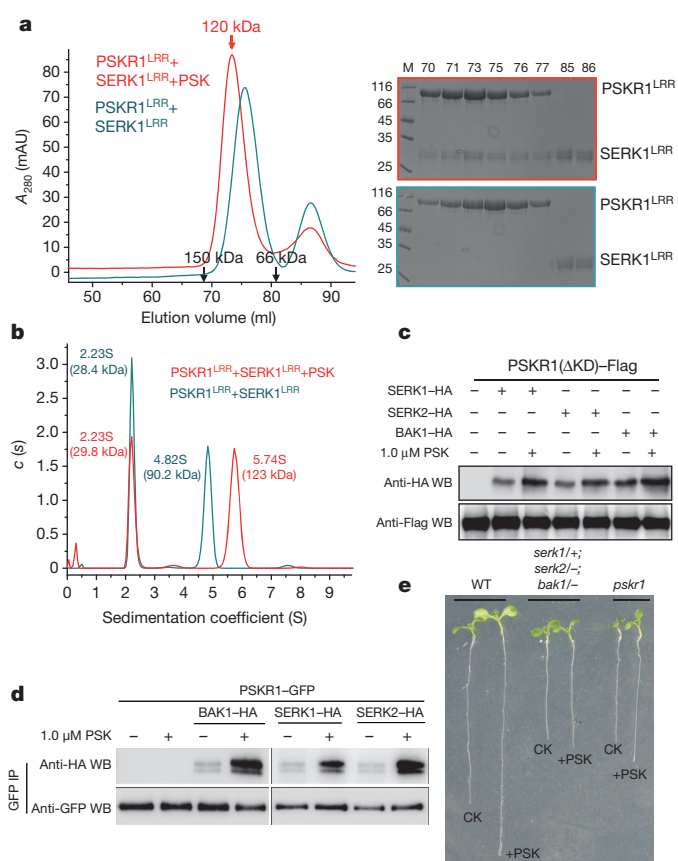
\*These authors contributed equally to this work.



**Figure 1 | Recognition mechanism of PSK by PSKR1<sup>LRR</sup>.** **a**, Overall structure of PSK-PSKR1<sup>LRR</sup> complex. Arrow indicates the position of PSK. ID, island domain. **b**, Detailed interactions of PSK (purple) with the ID (salmon) of PSKR1<sup>LRR</sup>. sY, sulfated tyrosine. **c**, Detailed interactions of PSK with the inner surface (cyan) of PSKR1<sup>LRR</sup>. **d**, Structural comparison of PSK-PSKR1<sup>LRR</sup> and PSK-DcPSKR<sup>LRR</sup>. **e**, Reducing PSK-PSKR1<sup>LRR</sup> interaction compromises PSKR1 to complement the shortened root phenotype of *pskr1-3* mutants. Average ( $\pm$ s.e.m.) primary root lengths of seedlings were determined in three independent experiments. Three independent overexpression lines (represented by -1, -2 and -3) per genotype were analysed ( $n = 30$  for each line, \* $P < 0.05$ , \*\*\* $P < 0.001$ , Student's  $t$ -test).

or DcPSKR<sup>LRR</sup>-SERK1/2<sup>LRR</sup> (Extended Data Fig. 4c, d) was heterodimeric in solution as indicated by gel filtration. Further supporting the gel filtration data, sedimentation-velocity analytical ultracentrifugation showed that PSKR1<sup>LRR</sup> formed a PSK-induced heterodimer with SERK1/2<sup>LRR</sup> or BAK1<sup>LRR</sup> (Fig. 2b and Extended Data Fig. 5).

Co-expression of full length Flag-conjugated PSKR1 (PSKR1-Flag) with haemagglutinin (HA)-conjugated SERKs resulted in rupture of *Arabidopsis* protoplasts quickly. We therefore used a KD truncated PSKR1 (PSKR1( $\Delta$ KD))-Flag and SERK1/SERK2/BAK1-HA for co-expression in protoplasts. Co-immunoprecipitation (Co-IP) assays showed that PSKR1( $\Delta$ KD) interacted with SERK1, SERK2 or BAK1 in protoplasts even in the absence of PSK (Fig. 2c), probably resulting from the endogenous PSK or their constitutive interaction, as observed for the BRI1-BAK1 interaction<sup>18</sup>. Importantly, the PSKR1( $\Delta$ KD)-SERK interactions were substantially increased in the PSK-treated protoplasts (Fig. 2c). Similar results were also obtained in *Arabidopsis* co-expressing PSKR1 and BAK1, SERK1 or SERK2 (Fig. 2d). Further supporting these results, the triple *serk1/+;serk2/-;bak1/-* mutant plants (where *serk1* is heterozygote *serk2* and *bak1* are

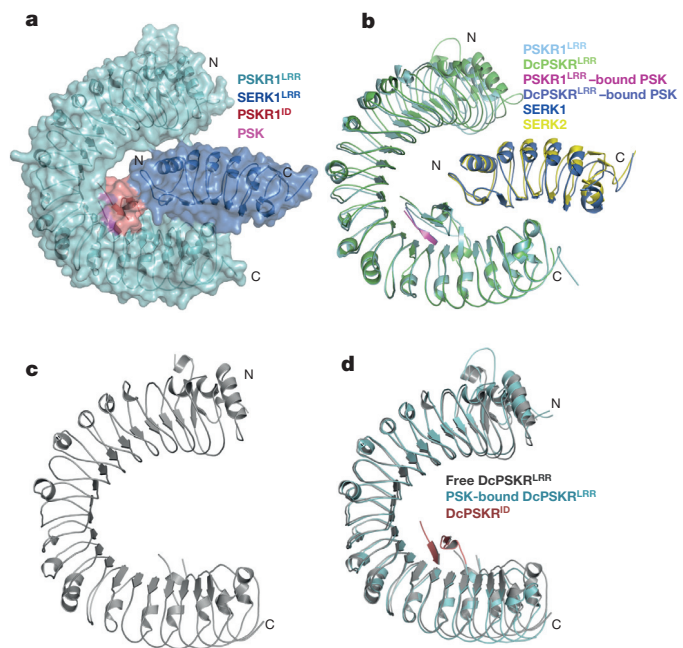


**Figure 2 | PSK promotes PSKR-SERK heterodimerization.** **a**, PSK induces PSKR1<sup>LRR</sup>-SERK1<sup>LRR</sup> heterodimerization. Left, superposition of the gel filtration chromatograms of the PSKR1<sup>LRR</sup> and SERK1<sup>LRR</sup> proteins. The red and black arrows indicate the elution positions of PSK-PSKR1<sup>LRR</sup>-SERK1<sup>LRR</sup> and molecular weight markers, respectively. mAU, micro-ultraviolet absorbance at 280 nm. Right, Coomassie blue staining of the peak fractions shown on the left following SDS-PAGE. M, molecular weight ladder (kDa). **b**, PSK induces a monomeric PSK-PSKR1<sup>LRR</sup>-SERK1<sup>LRR</sup> complex in sedimentation-velocity analytical ultracentrifugation. The peak sedimentation coefficients and the calculated molecular weights for the proteins indicated are shown. **c**, PSK promotes PSKR1-SERK interaction in *Arabidopsis* protoplasts. Flag-tagged PSKR1( $\Delta$ KD) and HA-tagged SERK1/2/BAK1 were co-expressed in WT *Arabidopsis* protoplasts, and their interactions were detected by co-immunoprecipitation (Co-IP). Each assay was repeated three times. Full blots are shown in Supplementary Data. **d**, PSK promotes PSKR1-SERK interaction *in planta*. Crude protein extracts from the treated and untreated plants overexpressing green fluorescent protein-conjugated PSKR1 (PSKR1-GFP) and SERK1/2/BAK1-HA were used for Co-IP experiments. Each assay was repeated three times. Full blots are shown in Supplementary Data. **e**, The *serk1/+;serk2/-;bak1/-* triple mutants are less sensitive to PSK in root growth. Wild-type or mutant *Arabidopsis* plants were grown for 10 days on plates with (+PSK) or without (CK) 1.0  $\mu$ M PSK. The image is representative of ten plants for each genotype.

homozygote) had shortened roots much less sensitive to PSK than the wild type (WT) plants, phenocopying the *pskr1-3* mutants (Fig. 2e). Only slightly shorter roots were observed in the single or double knockout plants (Extended Data Fig. 6a, b) that were still PSK-sensitive (Extended Data Fig. 6c), suggesting functional redundancy of SERKs in PSK-induced plant growth. It should be noted that the plant sensitivity to PSK was significantly reduced by inhibition of brassinosteroid-induced signalling<sup>19</sup> in which BAK1 and other SERK members play essential roles<sup>13</sup>.

We then solved the crystal structures of the PSK-PSKR1<sup>LRR</sup>-SERK1<sup>LRR</sup> (Fig. 3a and Extended Data Table 1) and PSK-DcPSKR<sup>LRR</sup>-SERK2<sup>LRR</sup> (Extended Data Fig. 7a and Extended Data Table 1) complexes. The structures of PSKR1<sup>LRR</sup> and DcPSKR<sup>LRR</sup>



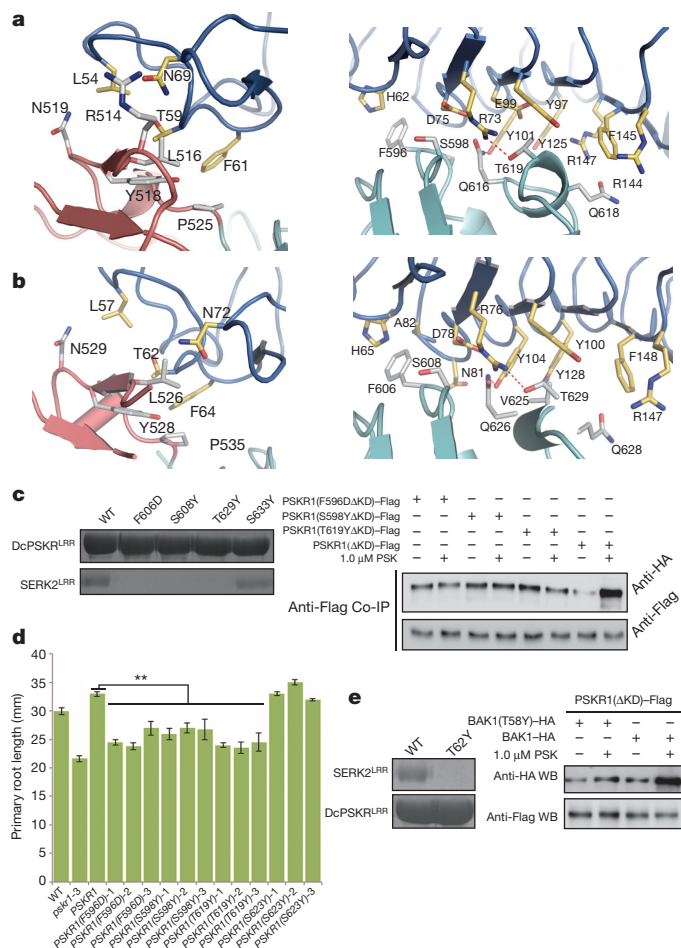


**Figure 3 | PSK stabilizes the PSKR<sup>ID</sup> for interaction with SERK<sup>LRR</sup>.** **a**, Overall structure of the PSK-PSKR<sup>LRR</sup>-SERK<sup>LRR</sup> complex. **b**, Structural comparison of PSK-PSKR<sup>LRR(31-638)</sup>-SERK<sup>LRR</sup> and PSK-DcPSKR<sup>LRR(29-647)</sup>-SERK<sup>LRR</sup>. **c**, Overall structure of the free DcPSKR<sup>LRR</sup>. **d**, PSK binding stabilizes the DcPSKR<sup>ID</sup>. Shown is the structural alignment of free DcPSKR<sup>LRR</sup> (residues 29–643, grey) and PSK-bound DcPSKR<sup>LRR</sup> (residues 29–643, cyan) with a r.m.s.d. of 0.66 Å.

are homologous (C $\alpha$  r.m.s.d. 1.49 Å over 600 amino acids) and an equivalent surface area is buried by their interaction with SERK<sup>LRR</sup> (984 Å<sup>2</sup>) and SERK<sup>LRR</sup> (973 Å<sup>2</sup>), respectively (Fig. 3b). SERK<sup>LRR</sup> binds the carboxy-terminal side of PSKR<sup>LRR</sup>, whereas PSKR<sup>ID</sup> contacts the amino-terminal side of SERK<sup>LRR</sup> (Fig. 3a). The structures of the two complexes are well aligned with that of the BRI1- but not the FLS2-containing complex (Extended Data Fig. 7b, c). Unlike the flg22- and brassinosteroid-mediated complexes<sup>14,20,21</sup>, PSK is not directly involved in the PSKR<sup>LRR</sup>-SERK<sup>LRR</sup> or DcPSKR<sup>LRR</sup>-SERK<sup>LRR</sup> interfaces (Fig. 3b). This is seemingly inconsistent with the PSK-promoted PSKR-SERK interaction. The structure of a free DcPSKR<sup>LRR</sup> (Extended Data Table 1) revealed that its ID is completely disordered (Fig. 3c and Extended Data Fig. 7d), sharply contrasting with the well-defined ID in PSK-bound DcPSKR<sup>LRR</sup> (Fig. 3d) or PSKR<sup>LRR</sup> (Fig. 1a). This demonstrates that PSK allosterically induces PSKR<sup>LRR</sup>-SERK<sup>LRR</sup> interaction.

PSKR<sup>ID</sup> interaction with the N-terminal side of SERK<sup>LRR</sup> is mainly mediated by van der Waals contacts (Fig. 4a). Centred at this interface is SERK<sup>Thr59</sup> that tightly packs against PSKR<sup>Leu516</sup> and PSKR<sup>Tyr518</sup>. Stacking of SERK<sup>Phe61</sup> against PSKR<sup>Pro525</sup> further fortifies the interactions around this interface (Fig. 4a, left panel). More extensive PSKR<sup>LRR</sup>-SERK<sup>LRR</sup> interactions come from contacts of the residues PSKR<sup>Phe596</sup>, PSKR<sup>Ser598</sup> and PSKR<sup>Thr619</sup> from one lateral side of PSKR<sup>LRR</sup> with the inner surface of SERK<sup>LRR</sup> (Fig. 4a, right panel). The PSKR<sup>LRR</sup>-SERK<sup>LRR</sup> interactions are highly conserved in the PSK-DcPSKR<sup>LRR</sup>-SERK<sup>LRR</sup> complex (Fig. 4b and Extended Data Fig. 1d).

DcPSKR(S608Y) and DcPSKR(T629Y), predicted to generate steric clashes with SERK<sup>LRR</sup> (Fig. 4b, right panel), led to loss of PSK-induced DcPSKR<sup>LRR</sup>-SERK<sup>LRR</sup> interaction (Fig. 4c, left panel and Extended Data Fig. 8a, b). A similar observation was also made for DcPSKR(F606D). Consistently, mutations of the equivalent residues PSKR<sup>Phe596</sup>, PSKR<sup>Ser598</sup> and PSKR<sup>Thr619</sup> (Fig. 4a) resulted in much less responsiveness to PSK for interaction with BAK1 in *Arabidopsis* protoplasts (Fig. 4c, right panel). Furthermore, mutations of these



**Figure 4 | Mutagenesis analysis of PSKR-SERK interaction.** **a**, Detailed interactions of the ID (salmon, left) and the C-terminal side (cyan, right) of PSKR<sup>LRR</sup> with SERK<sup>LRR</sup> (blue). **b**, Detailed interactions of the ID (salmon, left) and the C-terminal side (cyan, right) of DcPSKR<sup>LRR</sup> with SERK<sup>LRR</sup> (blue). **c**, Mutagenesis analysis of PSKR-SERK interaction. Left, interaction between WT or mutant DcPSKR<sup>LRR</sup> and SERK<sup>LRR</sup> in the presence of PSK as assayed in Fig. 2a. Right, mutations of critical PSKR1 residues render PSKR1(ΔKD)-BAK1 interaction less sensitive to PSK in *Arabidopsis* protoplasts as assayed in Fig. 2c. Full blots are shown in Supplementary Data. **d**, Reducing PSKR-SERK interaction compromises PSKR1 to complement the shortened root phenotypes of *pskr1-3* mutants. Average (±s.e.m.) primary root lengths of seedlings were determined in three independent experiments for each line ( $n = 30$ ,  $**P < 0.01$ , Student's  $t$ -test). **e**, Mutagenesis analysis of PSKR-SERK<sup>LRR</sup>/BAK1<sup>T58</sup> interaction. Left, SERK2(T62Y) disrupted PSK-induced DcPSKR<sup>LRR</sup>-SERK<sup>LRR</sup> interaction in solution as assayed in Fig. 2a. Right, BAK1(T58Y) is less sensitive to PSK for interaction with PSKR1(ΔKD) as assayed in Fig. 2c. Full blots are shown in Supplementary Data.

PSKR1 residues but not the controls (PSKR1(S623Y) or DcPSKR(S633Y)) (Fig. 4c and Extended Data Fig. 8a, b) reduced the ability of PSKR1 to complement the shorter roots of *pskr1-3* mutants and responsiveness of *Arabidopsis* plants to PSK (Fig. 4d and Extended Data Fig. 2c). SERK2(T62Y) is expected to generate similar effects on PSK-promoted PSKR-SERK interaction. Indeed, the SERK<sup>LRR</sup> mutant protein failed to form a PSK-induced complex with DcPSKR<sup>LRR</sup> (Fig. 4e, left panel and Extended Data Fig. 8a, b). Consistently, mutation of the equivalent residue BAK1<sup>Thr58</sup> (Extended Data Fig. 8c) rendered BAK1-PSKR1(ΔKD) interaction less responsive to PSK than wild-type BAK1 (Fig. 4e, right panel).

Our current study offers evidence that PSK promotes PSKR-SERK heterodimerization, providing a link between PSK perception and early intracellular signalling and further supporting the dimerization model<sup>14</sup>. Similar to brassinosteroid signalling<sup>22</sup>, PSK signalling also

negatively regulates pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI)<sup>23,24</sup>. However, expressions of disease-related genes were pathogen-induced in the *psk1-3* mutant plants<sup>23</sup>, similar to the *brl1* mutants<sup>22</sup>, whereas the *bak1* or *bak1 bkk1* (*serk4*) mutants displayed constitutive immune responses even under sterile growing conditions<sup>25,26</sup>. Thus, the roles played by SERK members in plant growth and disease resistance seem to be uncoupled, similar to those of BAK1 in brassinosteroid and PTI signalling<sup>27</sup>. These results can be reconciled by a previous model<sup>25,26</sup> postulating that SERK members negatively regulate a brassinosteroid-independent cell-death pathway induced by pathogens, which can be antagonized by the PSK signalling. PSK-enhanced PSKR–SERK heterodimerization can lead to transphosphorylation of the two RKs. Indeed, kinase activity of PSKR1 is essential for PSK-induced plant growth in *Arabidopsis*<sup>12</sup>.

Unlike flg22 and brassinosteroid, which mediate interactions between two LRR-RKs<sup>14,20,21</sup> by acting as ‘molecular glue’, PSK functions to stabilize the PSKR<sup>ID</sup>, which in turn recruits a SERK member to form a stable PSKR–SERK complex, resulting in allosteric activation of PSKR. The PSKR<sup>ID</sup> is shorter than that of BRI1, which is well structured even in the absence of ligand<sup>128,29</sup>. It therefore seems that ligand binding is required to complete the PSKR<sup>ID</sup>. Indeed, structural comparison showed that the PSKR<sup>ID</sup> together with PSK is similarly positioned to BRI1<sup>ID</sup> (Extended Data Fig. 7b). It will be interesting to investigate whether RLPs and some other RKs that contain an ID with a similar size and position (relative to the last LRR) to that of PSKR<sup>30</sup> use this mechanism for interaction with their partners.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 March; accepted 24 June 2015.

Published online 26 August 2015.

- Murphy, E., Smith, S. & De Smet, I. Small signaling peptides in *Arabidopsis* development: how cells communicate over a short distance. *Plant Cell* **24**, 3198–3217 (2012).
- Matsubayashi, Y. Posttranslationally modified small-peptide signals in plants. *Annu. Rev. Plant Biol.* **65**, 385–413 (2014).
- Matsubayashi, Y., Ogawa, M., Morita, A. & Sakagami, Y. An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. *Science* **296**, 1470–1472 (2002).
- Matsubayashi, Y., Ogawa, M., Kihara, H., Niwa, M. & Sakagami, Y. Disruption and overexpression of *Arabidopsis* phytosulfokine receptor gene affects cellular longevity and potential for growth. *Plant Physiol.* **142**, 45–53 (2006).
- Matsubayashi, Y. & Sakagami, Y. Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of *Asparagus officinalis* L. *Proc. Natl Acad. Sci. USA* **93**, 7623–7627 (1996).
- Srivastava, R., Liu, J. X. & Howell, S. H. Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in *Arabidopsis*. *Plant J.* **56**, 219–227 (2008).
- Komori, R., Amano, Y., Ogawa-Ohnishi, M. & Matsubayashi, Y. Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **106**, 15067–15072 (2009).
- Kutschmar, A. *et al.* PSK- $\alpha$  promotes root growth in *Arabidopsis*. *New Phytol.* **181**, 820–831 (2009).
- Stührwaldt, N., Dahlke, R. I., Steffens, B., Johnson, A. & Sauter, M. Phytosulfokine- $\alpha$  controls hypocotyl length and cell expansion in *Arabidopsis thaliana* through phytosulfokine receptor 1. *PLoS ONE* **6**, e21054 (2011).
- Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M. & Matsubayashi, Y. Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **104**, 18333–18338 (2007).
- Morillo, S. A. & Tax, F. E. Functional analysis of receptor-like kinases in monocots and dicots. *Curr. Opin. Plant Biol.* **9**, 460–469 (2006).
- Hartmann, J., Fischer, C., Dietrich, P. & Sauter, M. Kinase activity and calmodulin binding are essential for growth signaling by the phytosulfokine receptor PSKR1. *Plant J.* **78**, 192–202 (2014).
- Belkadir, Y., Yang, L., Hetzel, J., Dangl, J. L. & Chory, J. The growth-defense pivot: crisis management in plants mediated by LRR-RK surface receptors. *Trends Biochem. Sci.* **39**, 447–456 (2014).
- Han, Z., Sun, Y. & Chai, J. Structural insight into the activation of plant receptor kinases. *Curr. Opin. Plant Biol.* **20**, 55–63 (2014).
- Chinchilla, D., Shan, L., He, P., de Vries, S. & Kemmerling, B. One for all: the receptor-associated kinase BAK1. *Trends Plant Sci.* **14**, 535–541 (2009).
- Hanai, H. *et al.* A secreted peptide growth factor, phytosulfokine, acting as a stimulatory factor of carrot somatic embryo formation. *Plant Cell Physiol.* **41**, 27–32 (2000).
- Schmidt, E. D., Guzzo, F., Toonen, M. A. & de Vries, S. C. A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* **124**, 2049–2062 (1997).
- Wang, X. *et al.* Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Dev. Cell* **15**, 220–235 (2008).
- Heyman, J. *et al.* ERF115 controls root quiescent center cell division and stem cell replenishment. *Science* **342**, 860–863 (2013).
- Sun, Y. *et al.* Structural basis for flg22-induced activation of the *Arabidopsis* FLS2–BAK1 immune complex. *Science* **342**, 624–628 (2013).
- Santiago, J., Henzler, C. & Hothorn, M. Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. *Science* **341**, 889–892 (2013).
- Albrecht, C. *et al.* Brassinosteroids inhibit pathogen-associated molecular pattern-triggered immune signaling independent of the receptor kinase BAK1. *Proc. Natl Acad. Sci. USA* **109**, 303–308 (2012).
- Igarashi, D., Tsuda, K. & Katagiri, F. The peptide growth factor, phytosulfokine, attenuates pattern-triggered immunity. *Plant J.* **71**, 194–204 (2012).
- Mosher, S. *et al.* The tyrosine-sulfated peptide receptors PSKR1 and PSY1R modify the immunity of *Arabidopsis* to biotrophic and necrotrophic pathogens in an antagonistic manner. *Plant J.* **73**, 469–482 (2013).
- He, K. *et al.* BAK1 and BKK1 regulate brassinosteroid-dependent growth and brassinosteroid-independent cell-death pathways. *Curr. Biology* **17**, 1109–1115 (2007).
- Kemmerling, B. *et al.* The BRI1-associated kinase 1, BAK1, has a brassinolide-independent role in plant cell-death control. *Curr. Biology* **17**, 1116–1122 (2007).
- Schwessinger, B. *et al.* Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. *PLoS Genet.* **7**, e1002046, CrossRef (2011).
- She, J. *et al.* Structural insight into brassinosteroid perception by BRI1. *Nature* **474**, 472–476 (2011).
- Hothorn, M. *et al.* Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474**, 467–471 (2011).
- Fritz-Laylin, L. K., Krishnamurthy, N., Tor, M., Sjolander, K. V. & Jones, J. D. Phylogenomic analysis of the receptor-like proteins of rice and *Arabidopsis*. *Plant Physiol.* **138**, 611–623 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Huang and J. He for assistance with data collection, J. Li for the *serk1*, *serk2*, *bak1* single and triple mutant seeds and W. Li and W. Chu for providing facility support. This research was funded by Projects of International Cooperation and Exchanges NSFC (31420103906), Chinese Ministry of Science and Technology (2015CB910200) and State Key Program of National Natural Science of China (31130063) to J.C.; Chinese Natural Science Foundation (31330053) to W.Y. and Ministry of Science and Technology of China (2015CB910202) to H.L.

**Author Contributions** J.C., W.Y., J.W., H.L. and Z.H. designed the experiments. J.W., H.L., H.Z., T.W. and G.L. performed the experiments. Data were analysed by J.C., W.Y., J.W., H.L. and J.C.; J.C., W.Y., J.W., H.L. and Z.H. contributed to manuscript preparation. J.C. wrote the manuscript.

**Author Information** The atomic coordinates and structure factors have been deposited in the Protein Data Bank. The PDB code of free DcPSKR<sup>LRR</sup> is 4Z62. The PDB codes of PSK–PSKR1<sup>LRR</sup> and PSK–DcPSKR<sup>LRR</sup> are 4Z63 and 4Z5W, respectively. The PDB codes of PSK–PSKR1<sup>LRR</sup>–SERK1<sup>LRR</sup> and PSK–DcPSKR1<sup>LRR</sup>–SERK2<sup>LRR</sup> are 4Z64 and 4Z61, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C. ([chaijj@mail.tsinghua.edu.cn](mailto:chaijj@mail.tsinghua.edu.cn)) or W.Y. ([wcyang@genetics.ac.cn](mailto:wcyang@genetics.ac.cn)).



## METHODS

No statistical methods were used to predetermine sample size and the experiments were not randomized.

**Protein expression and purification.** The constructs of DcPSK<sup>LRR</sup> (residues 24–659), PSKR1<sup>LRR</sup> (residues 24–648), SERK1<sup>LRR</sup> (residues 1–213, N115D, N163Q), SERK2<sup>LRR</sup> (residues 1–216) and BAK1<sup>LRR</sup> (residues 1–220) with a C-terminal 6 × His tag were generated by standard PCR-based cloning strategy and their identities were confirmed by sequencing. DcPSK<sup>LRR</sup> and PSKR1<sup>LRR</sup> constructs were expressed in High Five insect cells at 22 °C using the pFastBac-1 vector (Invitrogen) with a modified N-terminal hemolysin signal peptide, and SERK1<sup>LRR</sup>, SERK2<sup>LRR</sup> and BAK1<sup>LRR</sup> constructs used the original pFastBac-1 vector. One litre of cells ( $2.0 \times 10^6$  cells ml<sup>-1</sup>) cultured in the medium from Expression Systems) was infected with 20 ml recombinant baculovirus and the media was harvested after 48 h. The proteins were purified using Ni-NTA (Novagen) and size-exclusion chromatography (Hiload200, GE Healthcare) in buffer containing 10 mM Bis-Tris pH 6.0 and 100 mM NaCl. The purified proteins were digested with endoglycosidase F1 and F3 at 18 °C overnight and further cleaned using gel filtration. The deglycosylated DcPSK<sup>LRR</sup> and PSKR1<sup>LRR</sup> proteins were concentrated to about 7.0 mg ml<sup>-1</sup> for crystallization. To crystallize the PSK–DcPSK<sup>LRR</sup>–SERK2<sup>LRR</sup> complex, the purified DcPSK<sup>LRR</sup>, SERK2<sup>LRR</sup> and the PSK peptide (synthesized by Silight Biotechnology, China) were mixed and incubated at 4 °C for 20 min. The mixture was subsequently subjected to gel filtration (Hiload200, GE Healthcare) in buffer containing 10 mM Bis-Tris pH 6.0, 100 mM NaCl. The purified complex was concentrated to about 7.0 mg ml<sup>-1</sup> for crystallization. Similar procedures were used for purification of the PSK–PSKR1<sup>LRR</sup>–SERK1<sup>LRR</sup> complex.

**Crystallization, data collection, structure determination and refinement.** Crystallization experiments were performed with hanging-drop vapour-diffusion methods by mixing equal volumes (1.0 µl) of protein and reservoir solution at 18 °C. Good quality crystals of DcPSK<sup>LRR</sup> were obtained in buffer containing 0.1 M Tris pH 8.5, 2.0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. For crystallization of PSK–DcPSK<sup>LRR</sup> or PSK–PSKR1<sup>LRR</sup> complex, a mixture of DcPSK<sup>LRR</sup> or PSKR1<sup>LRR</sup> and PSK peptide with a molar ratio of 1:5 was used for crystallization. Diffraction quality crystals of PSK–DcPSK<sup>LRR</sup> were obtained in buffer containing 0.3 M KH<sub>2</sub>PO<sub>4</sub>, 20% PEG(2,000) within 3 days, and for PSK–PSKR1<sup>LRR</sup>, good quality crystals appeared in buffer containing 0.1 M Bis-Tris pH 5.5, 2.0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> within 6 months. Diffraction quality crystals of the PSK–DcPSK<sup>LRR</sup>–SERK2<sup>LRR</sup> complex were obtained in buffer containing 0.1 M sodium citrate pH 5.5, 0.4 M KCl, 30% v/v pentaerythritol propoxylate (5/4 PO/OH) within one week, and for PSK–PSKR1<sup>LRR</sup>–SERK1<sup>LRR</sup>, high quality crystals emerged in buffer containing 0.1 M sodium acetate pH 4.5, 2.0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> over 6 months. All the diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF) on beam line BL17U1 using a CCD detector. The data were processed using HKL2000 (ref. 31). The crystal structure of PSK–DcPSK<sup>LRR</sup> was determined by molecular replacement (MR) with PHASER<sup>32</sup> using the structure of FLS2 (PDB code: 4MN8) as the initial searching model. The model from MR was built with the program COOT<sup>33</sup> and subsequently subjected to refinement by the program Phenix<sup>34</sup>. The other crystal structures were determined by MR using the structure of DcPSK<sup>LRR</sup> as the initial searching model. All the five crystal structures were refined by the program Phenix<sup>34</sup> with excellent stereochemistry (Extended Data Table 1). All the figures representing structures were prepared using PYMOL<sup>35</sup>.

**Microscale thermophoresis assay.** The microscale thermophoresis (MST) assay was performed as previously described<sup>36</sup>. The affinity of the purified DcPSK<sup>LRR</sup> (or its mutants) with PSK (or dPSK) was measured using the Monolith NT.115 from Nanotemper Technologies. Proteins were fluorescently labelled according to the manufacturer's protocol and the labelled protein used for each assay was about 200 nM. A solution of unlabelled peptide was diluted for appropriate serial concentration gradient. The samples were loaded into silica capillaries (Polymicro Technologies) after incubation at room temperature for 30 min. Measurements were performed at 20 °C in buffer containing 20 mM citric acid pH 5.0, 50 mM NaCl, and 0.05% Tween 20, by using 12% LED power and 40% MST power. The assays were repeated three times for each affinity measurement. Data analyses were performed using Nanotemper Analysis software and OriginPro 8.0 software provided by the manufacturer.

**Gel filtration assay.** The PSKR1<sup>LRR</sup> and SERK1<sup>LRR</sup> proteins purified as described above were subjected to gel filtration analysis (Hiload200, GE Healthcare) in the presence or absence of PSK. The PSKR1<sup>LRR</sup>, SERK1<sup>LRR</sup> proteins and PSK with a molar ratio of about 1:2:3 was mixed and incubated in 4 °C for 20 min before the gel filtration analysis in buffer containing 10 mM Bis-Tris pH 6.0, 100 mM NaCl. Samples from relevant fractions were applied to SDS–PAGE and visualized by Coomassie blue staining. Similar procedures were used for other interaction analysis of PSK<sup>LRR</sup>–SERK<sup>LRR</sup>. The DcPSK<sup>LRR</sup> and SERK2<sup>LRR</sup> mutants designed

to disrupt their interaction were also verified with the gel filtration assay described above.

**Sedimentation-velocity analytical ultracentrifugation.** Sedimentation velocity was performed with an XL-I analytical ultracentrifuge (Beckman Coulter) equipped with a four-cell An-60 Ti rotor for interaction analysis of PSKR1<sup>LRR</sup> and SERK2<sup>LRR</sup> in the presence or absence of PSK at 20 °C. For PSKR1<sup>LRR</sup> and SERK1<sup>LRR</sup> or BAK1<sup>LRR</sup>, an eight-cell An-50 Ti rotor was used. The molar ratio of PSKR1<sup>LRR</sup>, SERKs<sup>LRR</sup> proteins and PSK is about 1:2:3, and the total OD<sub>280</sub> is about 1.0. Buffer containing 10 mM Bis-Tris pH 6.0, 100 mM NaCl was used as the reference solution. All samples were applied at a speed of 45,000 rpm. Absorbance scans were taken at 280 nm at the intervals of 0.003 cm size in a radial direction. The different sedimentation coefficients, *c*(s), and molecular weight were calculated by SEDFIT V14.4f software.

**Plant materials and growth conditions.** *Arabidopsis thaliana* wild type Col-0 and *pskr1-3* (SALK\_008585) were obtained from *Arabidopsis* Biological Resource Center and reported to be a null mutant<sup>8</sup>. *serk1-8*, *serk2-1*, *bak1-4* and *serk1-8-/+ serk2-1-/- bak1-4-/-* triple mutant were generously provided by J. Li and each single mutant has been identified to be null<sup>37</sup>. Seeds were surface sterilized for 5 min in 20% NaClO<sub>3</sub> followed by 5 times of wash using sterile H<sub>2</sub>O and dispersed on 1/2 Murashige & Skoog (MS) media containing 1% agar and 10 g l<sup>-1</sup> sucrose, pH 5.8, in Petri dish. For PSK treatment of the seedlings, PSK was added in the MS media to different final concentrations. The sterilized seeds were vernalized for 3 days at 4 °C and grown for 10 days in normal condition (16 h of light/8 h of dark, 22–23 °C).

**Generation of constructs and plant transformation.** For stable transgenic plants, we generated the constructs of *PSKR1* coding sequences with different site mutations by subcloning the sequences into the pDONOR207 (Invitrogen) vector to the destination vector pWA43 or pWA53 by gateway recombination strategy (for PSKR1, DcPSK, F506A, R300A, W448A, T398L and D445A, the final target vector is pWA43; for F596D, S598Y, T619Y, S623Y and DcPSK<sup>ECD</sup>–PSKR1<sup>KD</sup>, the target vector is pWA53). pWA43 (hygromycin resistant in plants) and pWA53 (kanamycin resistant in plant) contained a CaMV 35S promoter driven C-terminal GFP coding sequence with the recombination sites in between and terminated by a 35S terminator. For the constructs used for transient protoplast transformation, the truncated PSKR1 coding sequence (PSKR1(ΔKD)) with the kinase domain deleted was fused with a C-terminal 3 × Flag affinity tag and inserted into the backbone of pBSK-35S: 35S Terminator after digestion with SmaI. For SERK1, SERK2, BAK1 and BAK1(T58Y) transient expression, the full-length coding sequences were inserted to pUC-SPYCE<sup>38</sup> which contains a C-terminal haemagglutinin affinity tag after digestion with SmaI. For co-expression *in planta*, SERK1, SERK2 and BAK1 inserted to PSPYCE-35S (kanamycin resistant in plant), which contains the same framework with pUC-SPYCE, were transformed to the T1 generation of pWA43-PSKR1 plants in the *pskr1-3* background. The transgenic plants were isolated by double selection on MS media containing kanamycin and hygromycin. *Arabidopsis* was transformed with these constructs by *Agrobacterium tumefaciens* (GV3101) by the floral dip method<sup>39</sup>.

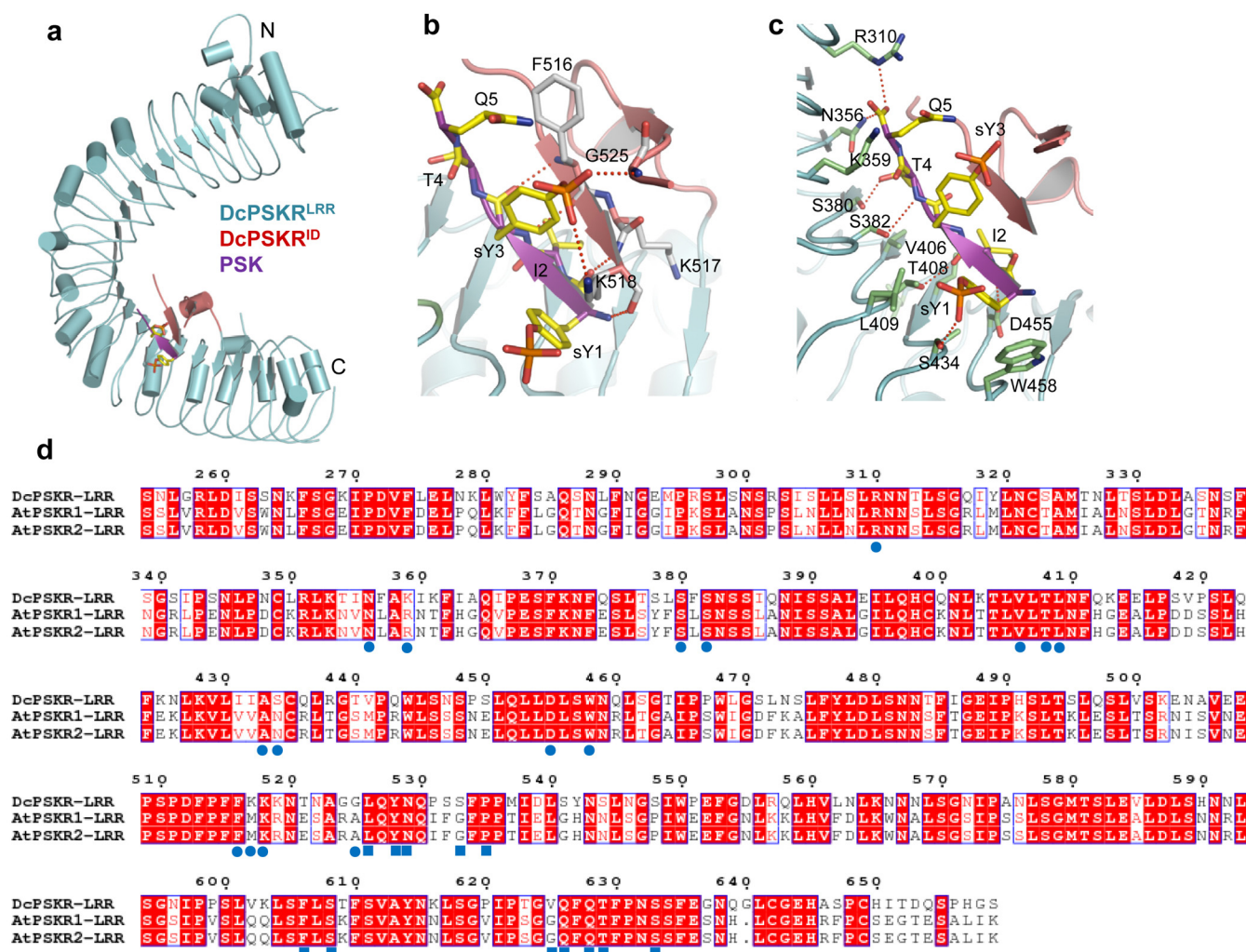
**Root length measurement and statistical analysis.** For each construct, ten transgenic overexpression lines in the *pskr1-3* mutant background were analysed and three lines representative for all lines were selected to present. 10-days seedlings grown in the greenhouse from the lines with *PSKR1* transcripts detected were subjected to primary root length measurement from photographs using Image J (National Institutes of Health, <http://rsb.info.nih.gov/ij>). To keep consistent seed fitness, only newly collected seeds at the same time were used for the assay. For each genotype, three independent experiments were performed. Student's *t*-test was performed to test statistical significance of means.

**PSK treatment and co-immunoprecipitation assay.** Protoplast transformation was performed according to the reported method<sup>40</sup> and cultured for 12 h at 22 °C. For each transformation, the culture of the transformed protoplasts was divided equally into two 50 ml centrifuge tubes. PSK peptide (diluted in H<sub>2</sub>O) was added to the final concentration of 1.0 µM in one tube and the same volume of H<sub>2</sub>O was added as mock treatment in the other tube. After 15 min of the treatment, the cells were harvested and lysed for 2 min in the lysis buffer (50 mM HEPES-KOH pH 7.5, 0.15 M KCl, 0.001 M EDTA, 0.1% Triton-X 100, 0.001 M DTT with freshly added proteinase inhibitor cocktail, (Roche)). The lysate was centrifuged at 10,000g for 10 min and the supernatant was subjected to coimmunoprecipitation (Co-IP) with agarose-conjugated anti-Flag antibody (Sigma-Aldrich, Cat. A220) for 3 h at 4 °C. The agarose beads were washed with the lysis buffer for 6 times, diluted in 1 × sample loading buffer and boiled for 5 min before SDS–PAGE. The following immunoblot was performed according to the standard procedure with anti-Flag (Sigma-Aldrich, Cat. F1804) and anti-HA antibody (Santa Cruz, Cat. sc-7392). For the Co-IP *in planta*, equal amounts of 14 days seedlings from the same transgenic lines (overexpressing PSKR1–GFP and SERK1/2/3–HA or PSKR1–GFP alone as a negative control) were treated on the MS media supplemented

with 1.0  $\mu$ M PSK and the MS media without PSK for 12 h. Then 5 g of the treated and untreated seedlings were collected and lysed for the following Co-IP experiments. GFP-trap agarose beads (ChromoTek, Cat. gta-200) were used for the affinity binding of the PSKR1–GFP fusion protein and anti-GFP-HRP (Miltenyi Biotec, Cat. 130-091-833) was used to detect the GFP epitope and anti-HA antibody for HA epitope. Each Co-IP experiment was repeated at least three times.

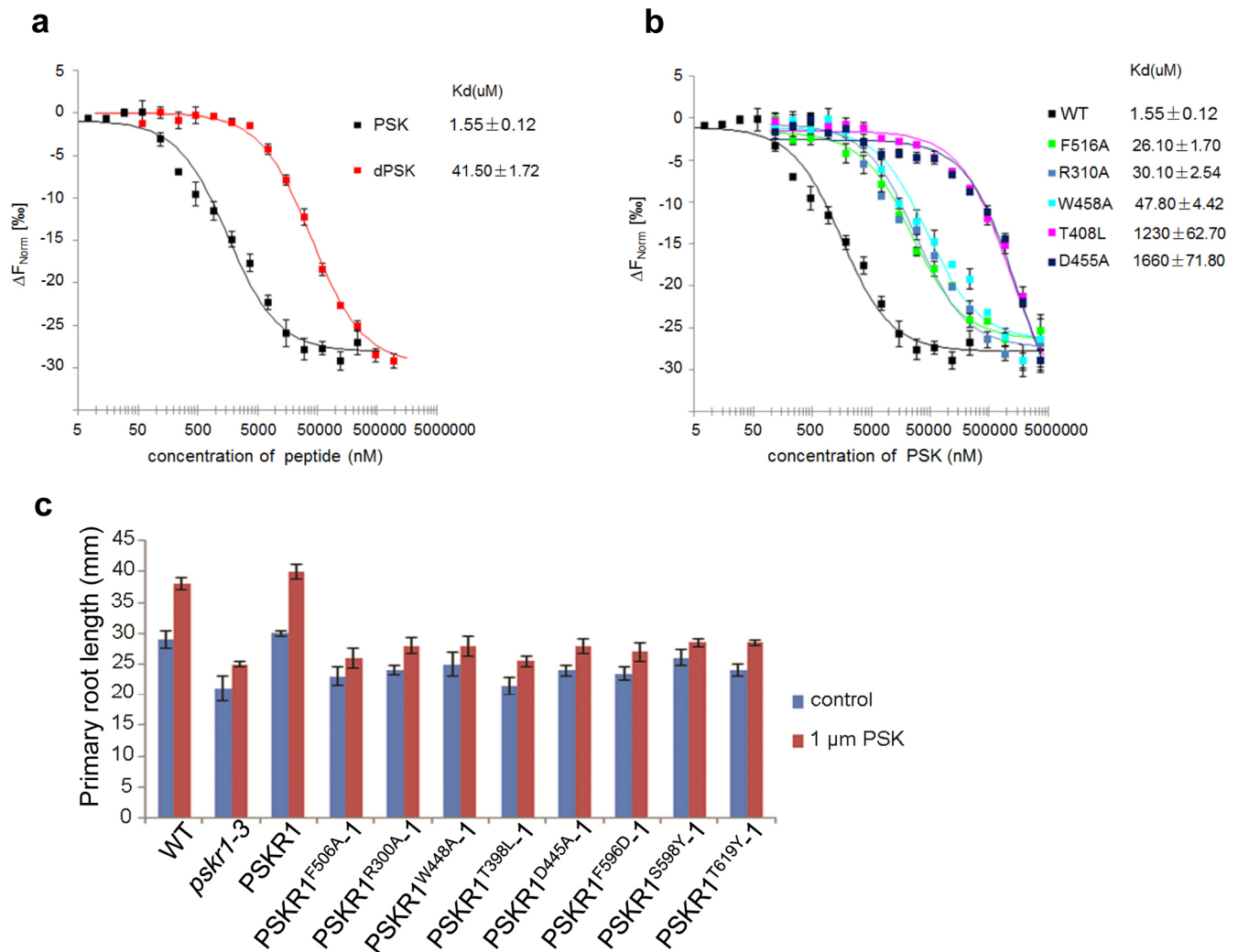
31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
35. DeLano, W. L. PyMOL Molecular Viewer (<http://www.pymol.org>) (2002).
36. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342–353 (2011).
37. Gou, X. *et al.* Genetic evidence for an indispensable role of somatic embryogenesis receptor kinases in brassinosteroid signaling. *PLoS Genet.* **8**, e1002452 (2012).
38. Walter, M. *et al.* Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438 (2004).
39. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
40. Yoo, S.-D., Cho, Y.-H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2**, 1565–1572 (2007).





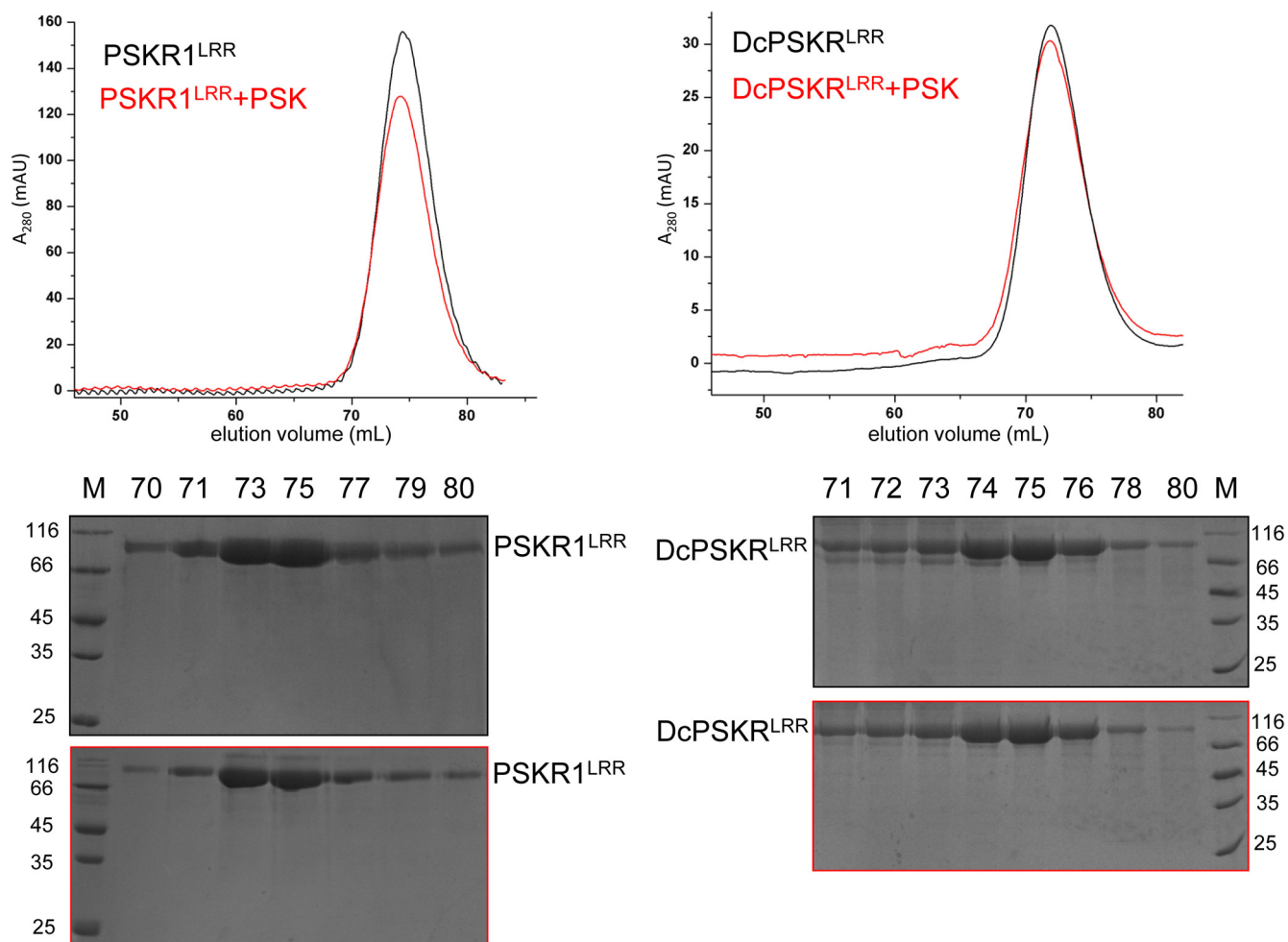
**Extended Data Figure 1 | Recognition mechanism of PSK by PSKR<sup>LRR</sup> is highly conserved.** **a**, Overall structure of PSK-DcPSKR<sup>LRR</sup> complex. The sulfated tyrosines of PSK are shown in stick. Colour codes are indicated. ID, island domain; N, N terminus; C, C terminus. **b**, Detailed interactions between PSK (purple) and the island domain (salmon) of DcPSKR<sup>LRR</sup>. Dashed lines indicate polar interactions. **c**, Detailed interactions between PSK and the inner

side (cyan) of DcPSKR<sup>LRR</sup>. **d**, PSKRs are conserved in PSK perception and interaction with SERKs. Sequence alignment of the ectodomains of carrot DcPSKR and *Arabidopsis* PSKR1/2. Conserved and similar residues are boxed with red ground and red font, respectively. Residues involved in recognition of PSK and interaction with a SERK member are indicated with blue solid circles and squares at the bottom, respectively.



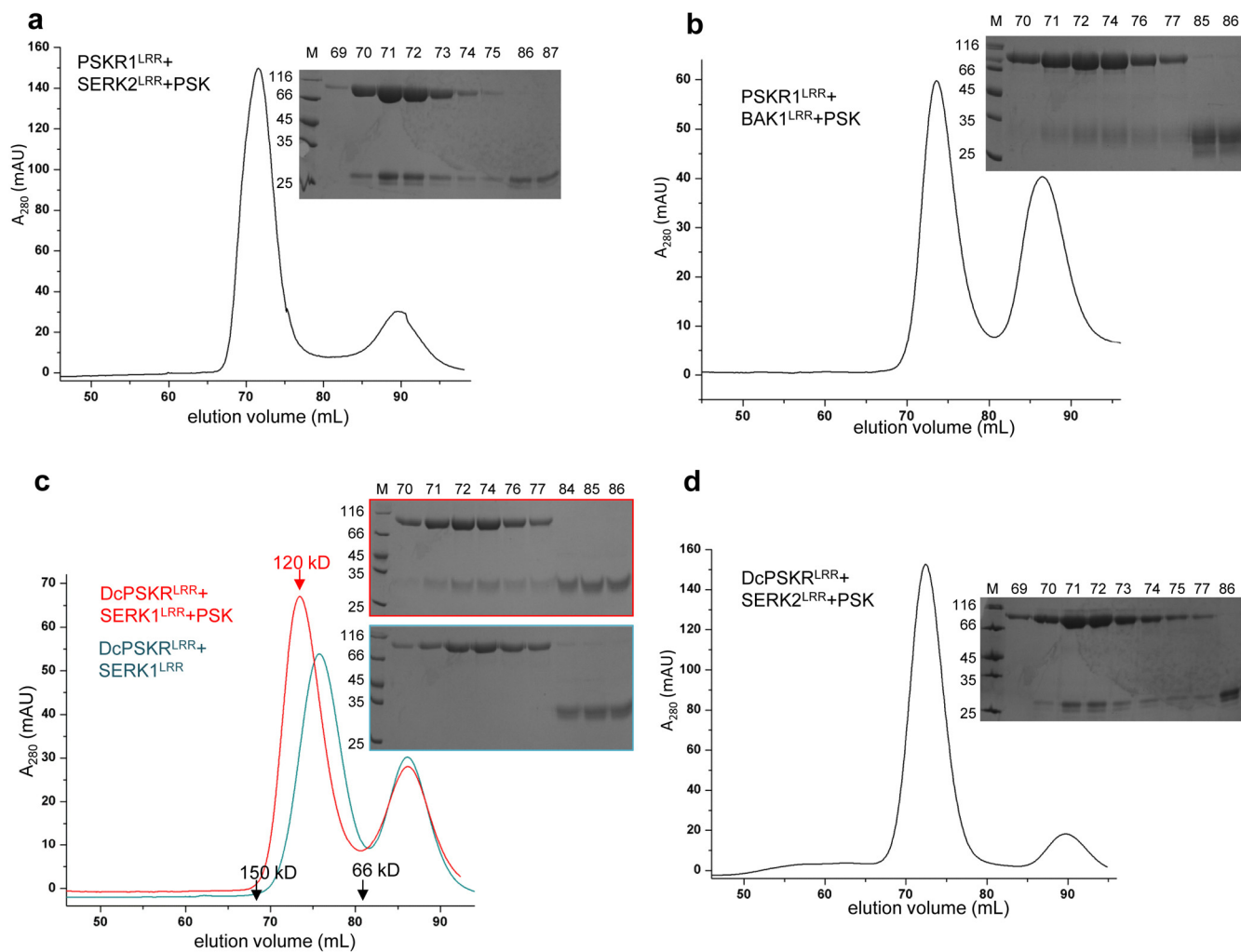
**Extended Data Figure 2 | Mutagenesis analysis of PSKR recognition of PSK and PSKR-SERK interaction.** **a**, Sulfation enhances PSK interaction with DcPSKR<sup>LRR</sup>. Quantification of binding affinity between DcPSKR<sup>LRR</sup> and PSK or the desulfated peptide (dPSK) by MST (MicroScale Thermophoresis). Data points indicate the difference in normalized fluorescence (%) generated by PSK or dPSK binding DcPSKR<sup>LRR</sup> protein, and curves indicate the calculated fits. Error bars represent standard error of 3 independent measurements. **b**, Mutagenesis analysis of DcPSKR<sup>LRR</sup> by MST. Quantification of binding

affinity between WT DcPSKR<sup>LRR</sup> or various mutants as indicated and PSK by MST. Error bars represent standard error of 3 independent measurements. **c**, *pskr1-3* plants transformed with mutated *PSKR1* which compromised PSK or SERKs binding are less responsive to PSK than wild type or *pskr1-3* transformed with *PSKR1*. The line was the same as that used in Fig. 1e and 4d. Average ( $\pm$ s.e.m.) primary root lengths of seedlings were determined in three independent experiments with 30 seedlings analysed per genotype in the presence or absence of 1.0  $\mu\text{M}$  PSK.



**Extended Data Figure 3 | PSK binding induces no oligomerization of PSKR<sup>LRR</sup>.** Shown on the top is superposition of the gel filtration chromatograms of the PSKR1<sup>LRR</sup> (left) or DcPSKR<sup>LRR</sup> (right) protein in the absence (grey) and presence (red) of PSK. The vertical and horizontal axes

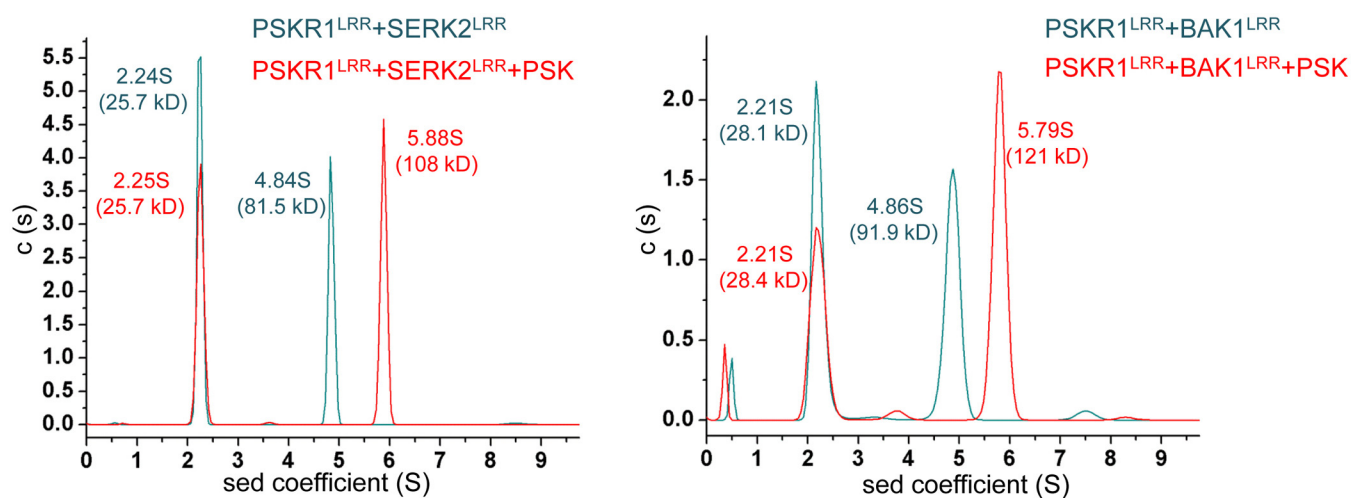
represent ultraviolet absorbance ( $\lambda = 280$  nm) and elution volume (ml), respectively. Bottom, Coomassie blue staining of the peak fractions shown on the top following SDS-PAGE. M, molecular weight ladder (kDa).



**Extended Data Figure 4 | PSK induces PSKR1<sup>LRR</sup> or DcPSKR<sup>LRR</sup> interaction with SERK members in gel filtration.** **a**, PSK induces PSKR1<sup>LRR</sup>-SERK2<sup>LRR</sup> heterodimerization. Right, Coomassie blue staining of the peak fractions shown on the left following SDS-PAGE. M, molecular weight ladder (kDa). **b**, PSK induces PSKR1<sup>LRR</sup> heterodimerization with BAK1<sup>LRR</sup>. The assay was performed as described in **a**. **c**, PSK induces DcPSKR<sup>LRR</sup>

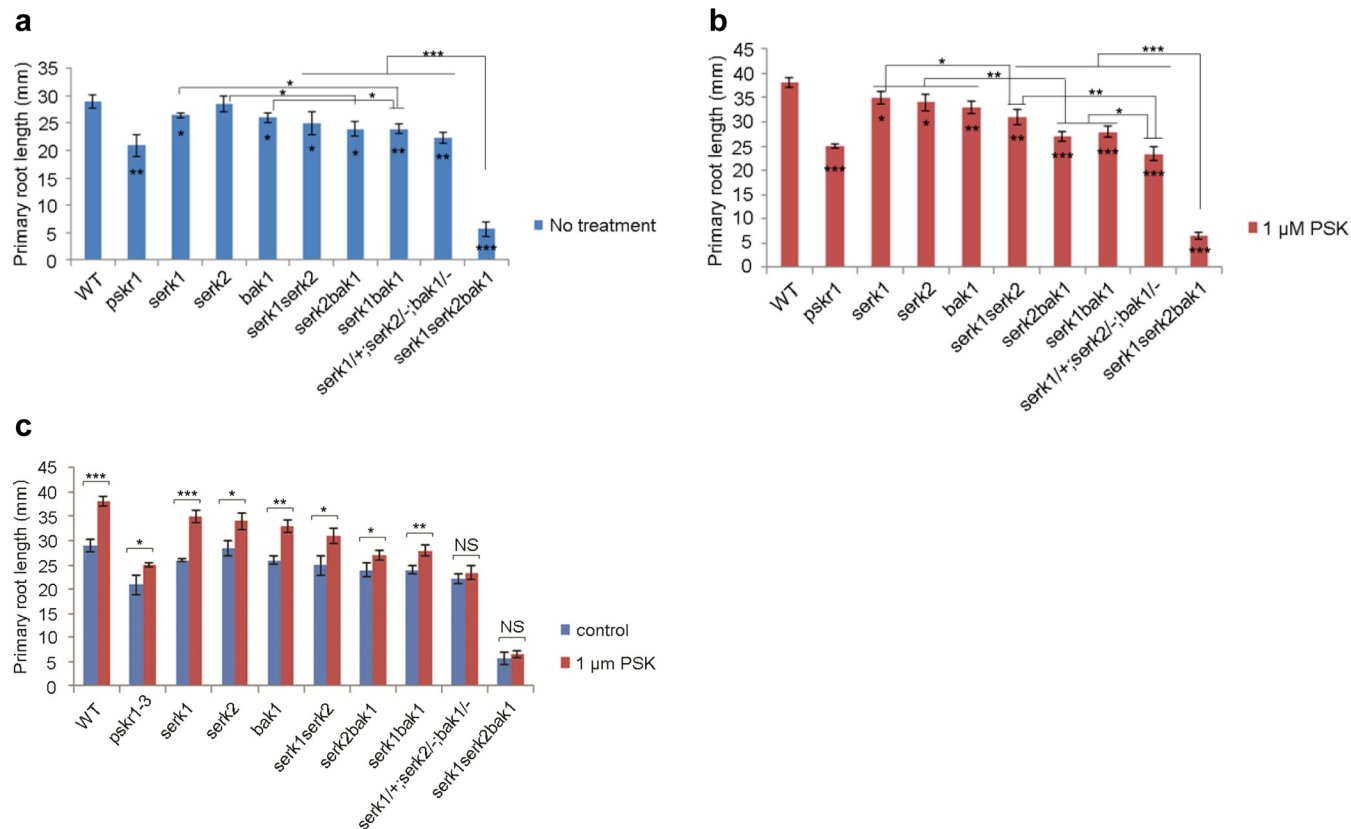
heterodimerization with SERK1<sup>LRR</sup>. The assays were performed as described in **a**. The red and black arrows indicate the elution position of PSK-DcPSKR<sup>LRR</sup>-SERK1<sup>LRR</sup> and the retention volumes of molecular weight markers, respectively. **d**, PSK induces DcPSKR<sup>LRR</sup> heterodimerization with SERK2<sup>LRR</sup>. The assay was performed as described in **a**.





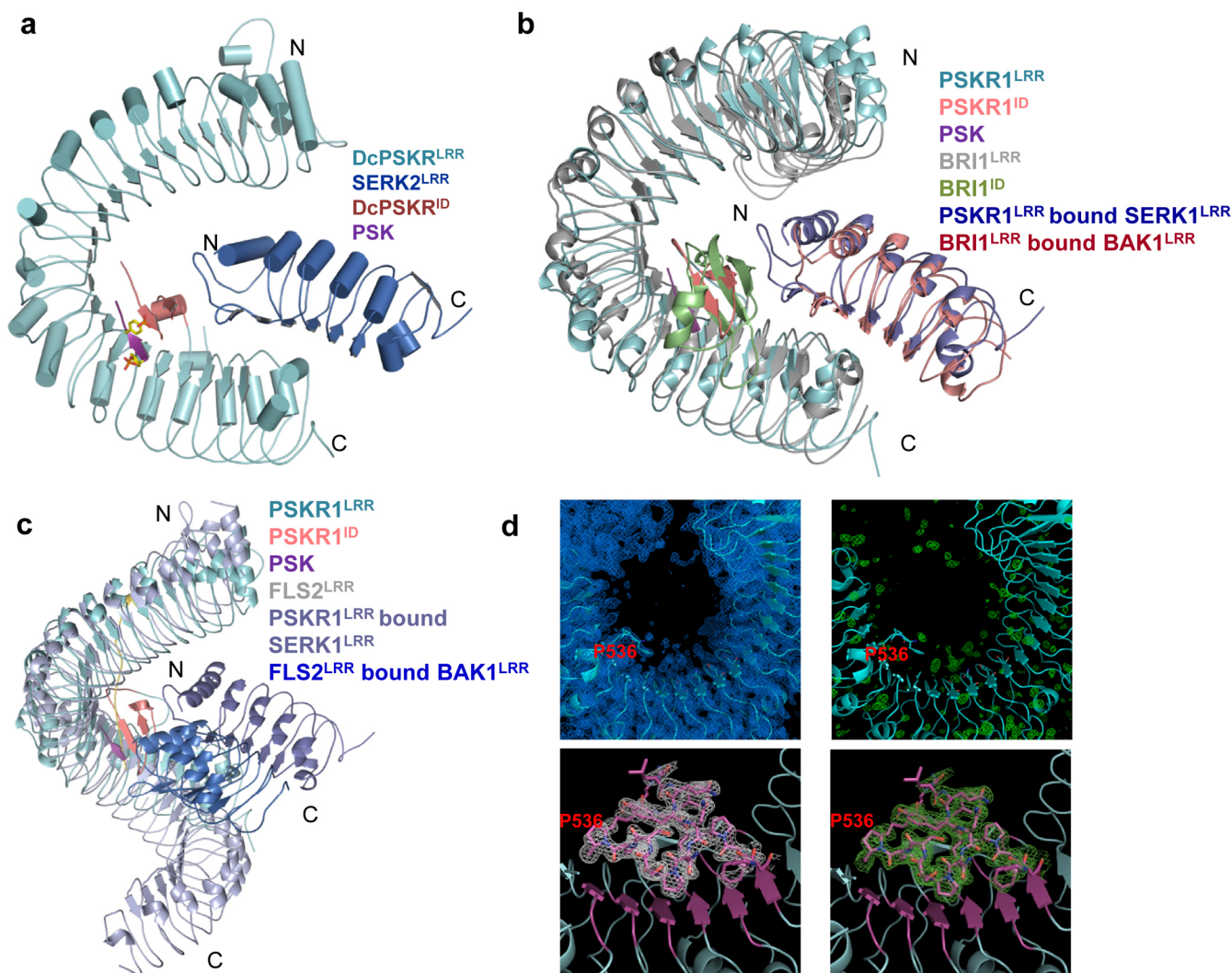
**Extended Data Figure 5 | PSK induces PSKR1<sup>LRR</sup> interaction with SERK members in sedimentation-velocity analytical ultracentrifugation.** PSK induces PSKR1<sup>LRR</sup>-SERK2<sup>LRR</sup> (left panel) or PSKR1<sup>LRR</sup>-BAK1<sup>LRR</sup> (right panel) interaction in sedimentation-velocity analytical ultracentrifugation assays. The assays were performed as described in Fig. 2b. The glycoprotein

nature of PSKR1<sup>LRR</sup> may confer to the slight difference of calculated molecular weights. PSK induced the formation of a monomeric PSK-PSKR1<sup>LRR</sup>-SERK2<sup>LRR</sup> or PSK-PSKR1<sup>LRR</sup>-BAK1<sup>LRR</sup> complex, leading to the shift of PSKR1<sup>LRR</sup> to a higher S.



**Extended Data Figure 6 | SERK members function redundantly in PSK-induced plant growth.** **a–c**, Average ( $\pm$ s.e.m.) primary root lengths of seedlings were determined for the wild-type or *SERK* knockout *Arabidopsis* plants grown for 10 days on plates with (red) or without (blue) 1.0  $\mu$ M PSK. Three independent experiments per genotype with 30 seedlings were performed. The statistics are shown in **a**, **b** and **c**. All the genotypes are compared in the absence of PSK in **a** and in the presence of PSK in **b**. The single

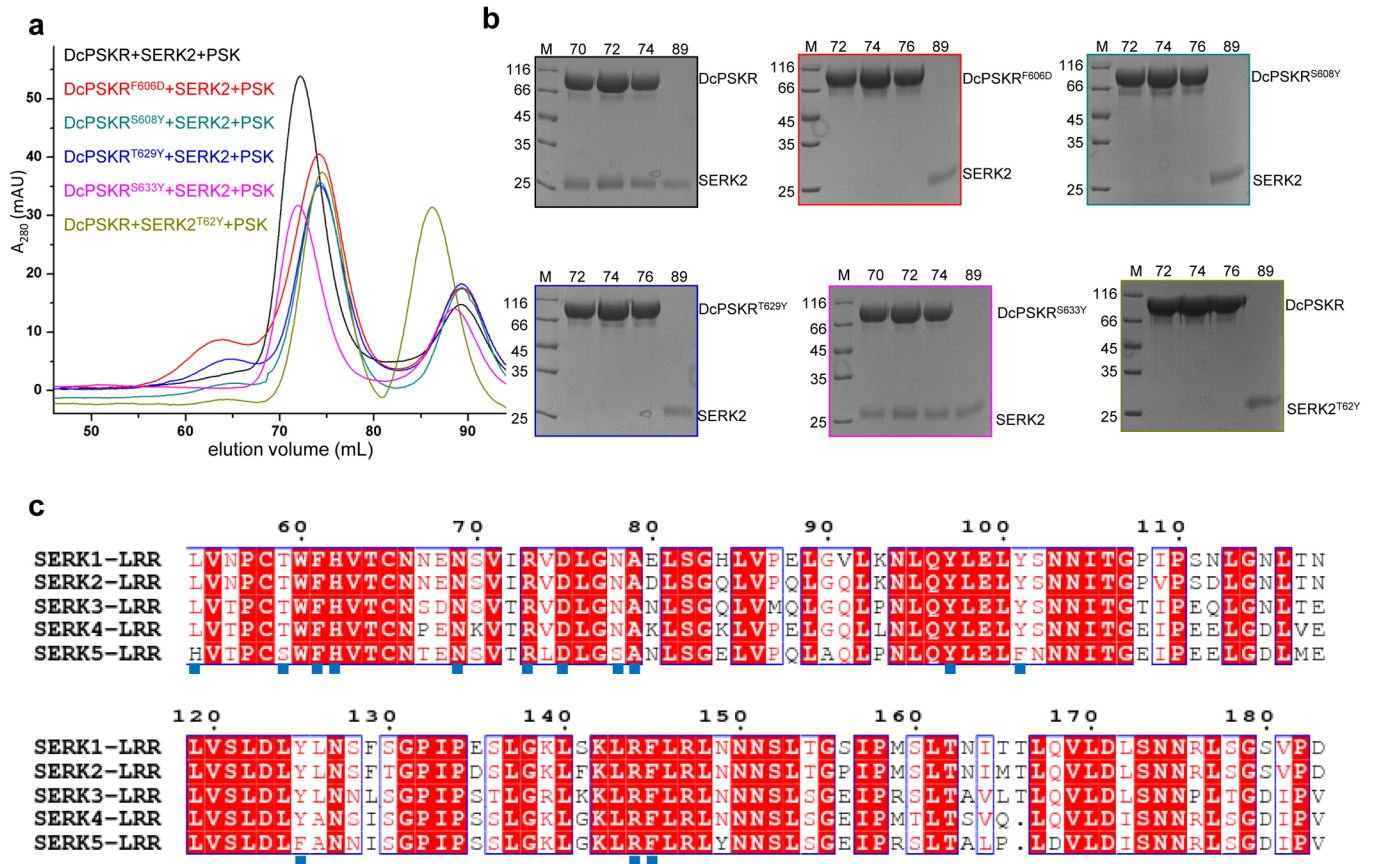
or double *SERK* knockout plants only showed slightly shortened roots compared to the triple mutants. Asterisks within the bars indicate significant difference between the wild type and *SERK* knockout mutants and those above the bars indicate significant difference between different *SERK* knockout mutants. Each genotype in the presence and absence of PSK is compared in **c**. Student's *t*-test, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ . NS, non-significant ( $P > 0.05$ ).



**Extended Data Figure 7 | Different mechanism of PSK induced PSKR-SERK interaction compared to BRI1-BAK1 or FLS2-BAK1 complex.**

**a**, Overall structure of PSK-DcPSKR<sup>LRR</sup>-SERK2<sup>LRR</sup> complex. **b**, Structural comparison of PSK-PSKR<sup>LRR</sup>-SERK1<sup>LRR</sup> and brassinosteroid-BRI1<sup>LRR</sup>-BAK1<sup>LRR</sup>. The structure of PSKR<sup>LRR</sup> (residues 77–634) was used as the template for alignment with that of BRI1 (residues 174–766; PDB code 4M7E) with a r.m.s.d. of 2.43 Å. **c**, Structural comparison of PSK-PSKR<sup>LRR</sup>-SERK1<sup>LRR</sup> and flg22-FLS2<sup>LRR</sup>-BAK1<sup>LRR</sup>. The structure of PSKR<sup>LRR</sup> (residues 82–554) was used as the template for alignment with that of FLS2 (residues 79–509; PDB code 4MN8) with a r.m.s.d. of 4.4 Å. SERK1<sup>LRR</sup> bound by PSKR<sup>LRR</sup> rotates about 30 degrees and shifts about 20 Å relative to the BAK1<sup>LRR</sup>-bound

FLS2<sup>LRR</sup>. **d**, Electron density around the island domain of DcPSKR<sup>LRR</sup> and PSK-bound DcPSKR<sup>LRR</sup> in the finally refined structures. Top panel, electron density  $2F_o - F_c$  (left) and  $F_o - F_c$  (right) contoured at 1.30 sigma and 2.7 sigma, respectively, for the finally refined free DcPSKR<sup>LRR</sup> structure. Bottom panel: electron density  $2F_o - F_c$  (left) and  $F_o - F_c$  (right) omitted around the island domain in the structure of PSK-bound DcPSKR<sup>LRR</sup>. The island domain (residues 511–535) and the  $\beta$ -sheet (residues 474–480, 450–456, 427–432, 402–408, 376–381 and 352–357) interacting with the ID were not included in refinement and electron density calculation. All the deleted residues are shown in pink. The marker residue proline 536 is shown in red.



**Extended Data Figure 8 | Mutagenesis analysis of DcPSKR<sup>LRR</sup>-SERK2<sup>LRR</sup> interaction.** **a**, Superposition of the gel filtration chromatograms of the mutant DcPSKR<sup>LRR</sup> and SERK2<sup>LRR</sup> proteins in the presence of PSK. The assays were performed as described in Extended Data Fig. 4a. **b**, Coomassie blue staining of the peak fractions shown on the left chromatograms following SDS-PAGE. M, molecular weight ladder (kDa). **c**, The amino acids of SERKs

involved in PSKR interaction are conserved. Sequence alignment of the ectodomains of SERK family proteins. Conserved and similar residues are boxed with red ground and red font, respectively. Residues involved in interaction with PSKR are indicated with blue solid squares at the bottom. The sequence of SERK3 is 100% identical to BAK1.



Extended Data Table 1 | Data collection and refinement statistics

Data set	DcPSK <sup>LRR</sup>	PSK-DcPSK <sup>LRR</sup>	PSK-DcPSK <sup>LRR</sup> - SERK2 <sup>LRR</sup>	PSK-PSK1 <sup>LRR</sup>	PSK-PSK1 <sup>LRR</sup> - SERK1 <sup>LRR</sup>
Wavelength (Å)	1.000	1.000	1.000	1.000	1.000
Resolution (Å)	99.0-2.9 (2.95-2.9)	99.0-2.2 (2.24-2.2)	99.0-2.75 (2.8-2.75)	99.0-2.5 (2.54-2.5)	99.0-2.65 (2.7-2.65)
Space group	C222 <sub>1</sub>	P1	C2	P4 <sub>3</sub> 22	C222 <sub>1</sub>
a, b, c (Å)	90.0, 98.8, 227.3	66.7, 75.7, 93.9	486.2, 73.5, 67.3	92.9, 92.9, 242.5	152.5, 220.9, 105.4
$\alpha, \beta, \gamma$ (°)	90.0, 90.0, 90.0	111.3, 105.7, 97.2	90.0, 95.8, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Unique reflections	21,448(1,019)	78,850(2,566)	54,320(2,762)	36,529(1,779)	49,329(2,490)
Completeness	93.6% (92.6%)	97.3% (87.0%)	88.8% (87.0%)	98.4% (99.6%)	96.8% (98.8%)
R <sub>sym</sub> (%)	12.6(47.4)	7.4(54.9)	7.9(40.9)	10.3(56.6)	10.0 (48.4)
redundancy	5.3(5.3)	3.9(3.9)	2.0(1.9)	5.3(5.7)	4.2(4.2)
<i>I</i> / $\sigma$	18.5(6.4)	17.9(5.9)	14.7(3.1)	26.0(4.4)	15.5(2.65)
<b>Statistics for refinement</b>					
Resolution (Å)	99-2.9 (3.03-2.9)	99.0-2.2 (2.23-2.20)	99.0-2.75 (2.8-2.75)	99.0-2.5 (2.58-2.5)	99.0-2.66 (2.71-2.66)
No. of RFs	21,401(2,391)	78,850(2,566)	54,303(2,613)	36,369(2,715)	49,253(2,282)
Completeness	93.4%	97.1%	87.7%	98.1%	95.7%
R <sub>work</sub> /R <sub>free</sub>	20.2 (24.9)/	21.2(29.0)/	20.0(24.2)/	22.9(26.7)/	20.0(28.6) /
(%)	25.4(32.6)	26.2(33.7)	26.2(38.3)	27.6(35.1)	24.6(36.7)
R.m.s.d					
Bond (degree)	1.418	1.389	1.445	1.481	1.265
length (Å)	0.008	0.008	0.008	0.008	0.008
Ramachandran	Favored: 96.3%	Favored: 98.1%	Favored: 96.6%	Favored: 96.6%	Favored: 89.6%
Plot	Allowed: 3.7%	Allowed: 1.8%	Allowed: 3.3%	Allowed: 3.3%	Allowed: 9.4%
	Outliers: 0.0%	Outliers: 0.1%	Outliers: 0.1%	Outliers: 0.1%	Outliers: 1.0%

RF, reflection.  $R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \bar{I}_h| / \sum_h \sum_i I_{h,i}$ , where  $\bar{I}_h$  is the mean intensity of the *i* observations of symmetry related reflections of *h*.  $R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ , where  $F_{\text{obs}} = F_p$ , and  $F_{\text{calc}}$  is the calculated protein structure factor from the atomic model. R.m.s.d. in bond lengths and angles are the deviations from ideal values.

# Structural basis of JAZ repression of MYC transcription factors in jasmonate signalling

Feng Zhang<sup>1,2,3\*</sup>, Jian Yao<sup>2,4\*</sup>, Jiyuan Ke<sup>1\*</sup>, Li Zhang<sup>2,5</sup>, Vinh Q. Lam<sup>6</sup>, Xiu-Fang Xin<sup>2</sup>, X. Edward Zhou<sup>1</sup>, Jian Chen<sup>1,7</sup>, Joseph Brunzelle<sup>8</sup>, Patrick R. Griffin<sup>6</sup>, Mingguo Zhou<sup>3</sup>, H. Eric Xu<sup>1,9</sup>, Karsten Melcher<sup>1</sup> & Sheng Yang He<sup>2,5,10</sup>

The plant hormone jasmonate plays crucial roles in regulating plant responses to herbivorous insects and microbial pathogens and is an important regulator of plant growth and development<sup>1–7</sup>. Key mediators of jasmonate signalling include MYC transcription factors, which are repressed by jasmonate ZIM-domain (JAZ) transcriptional repressors in the resting state. In the presence of active jasmonate, JAZ proteins function as jasmonate co-receptors by forming a hormone-dependent complex with COI1, the F-box subunit of an SCF-type ubiquitin E3 ligase<sup>8–11</sup>. The hormone-dependent formation of the COI1–JAZ co-receptor complex leads to ubiquitination and proteasome-dependent degradation of JAZ repressors and release of MYC proteins from transcriptional repression<sup>3,10,12</sup>. The mechanism by which JAZ proteins repress MYC transcription factors and how JAZ proteins switch between the repressor function in the absence of hormone and the co-receptor function in the presence of hormone remain enigmatic. Here we show that *Arabidopsis* MYC3 undergoes pronounced conformational changes when bound to the conserved Jas motif of the JAZ9 repressor. The Jas motif, previously shown to bind to hormone as a partly unwound helix, forms a complete  $\alpha$ -helix that displaces the amino (N)-terminal helix of MYC3 and becomes an integral part of the MYC N-terminal fold. In this position, the Jas helix competitively inhibits MYC3 interaction with the MED25 subunit of the transcriptional Mediator complex. Our structural and functional studies elucidate a dynamic molecular switch mechanism that governs the repression and activation of a major plant hormone pathway.

To understand the structural basis of the interactions between MYC transcription factors and JAZ repressors, we first used yeast two-hybrid assays to determine the JAZ-binding regions within MYC2, MYC3 and MYC4. A conserved region of ~200 amino acids (amino acids 55–259, 44–234 and 55–253 in MYC2, MYC3 and MYC4, respectively) within the N termini of all three proteins that encompasses the previously defined JAZ-interacting domain (JID)<sup>13,14</sup> and the transcription activation domain (TAD)<sup>13,15</sup> was sufficient to interact with JAZ9 (Extended Data Figs 1a and 2a). Similarly, we identified a region of 17 amino acids within the Jas motif of JAZ9 (polyA-Jas) that is required and sufficient to interact with MYC3 (Extended Data Fig. 1b). Interestingly, this Jas motif shares the same segment of JAZ proteins that interacts with COI1 (ref. 16), but is four amino acids shorter at the N terminus (Extended Data Fig. 1c). We confirmed these results using AlphaScreen luminescence proximity assays with His6-tagged MYC proteins and biotinylated JAZ8, JAZ9 and JAZ12 peptides (Extended Data Figs 1d and 2b).

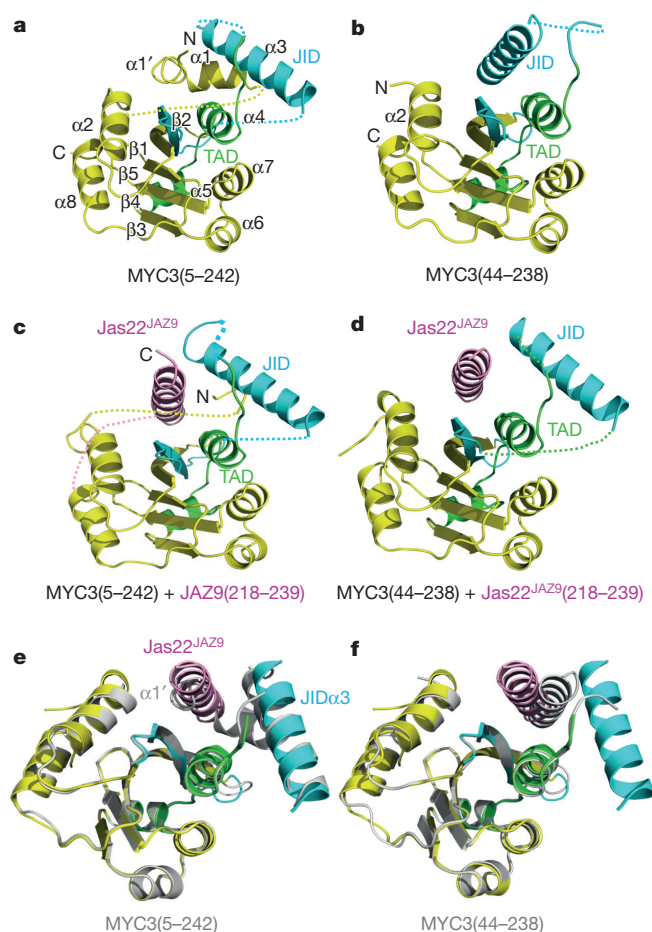
On the basis of our mapping results, we generated 15 MYC2/3/4 N-terminal truncated proteins of various lengths (Extended Data

Figs 1d and 2b). MYC3(44–238) and MYC3(5–242) yielded high-quality crystals that diffracted X-rays to 2.2 Å and 2.1 Å resolution, respectively (Extended Data Table 1). We solved the structure of selenomethionine-modified MYC3(44–238) by selenium single-wavelength anomalous diffraction (Se-SAD) and the structure of MYC3(5–242) by molecular replacement using the structure of MYC3(44–238) as search model (Fig. 1a, b and Extended Data Fig. 3). The proteins formed a helix–sheet–helix sandwich fold, in which eight  $\alpha$ -helices are wrapped around a central five-stranded anti-parallel  $\beta$ -sheet (Fig. 1a). Remarkably, while a hallmark of acidic TAD is that they are unstructured when not bound to a target in the transcriptional machinery<sup>17–19</sup>, the MYC3 TAD is well resolved and forms a loop–helix–loop–helix motif that packs against the JID with the N-terminal TAD helix and against  $\beta$ -strands 3–5 with the carboxy (C)-terminal TAD helix (Fig. 1a, b and Extended Data Fig. 3). To our knowledge, this is the first example in which a non-complexed acidic TAD has a well-resolved structure. The JID consists of the top ( $\beta$ 2) strand of the  $\beta$ -sheet, the long  $\alpha$ 3-helix and two unresolved linkers (Fig. 1a, b and Extended Data Fig. 3a). In MYC3(5–242), the JID forms together with the  $\alpha$ 4-helix of the TAD groove. The N-terminal MYC helix ( $\alpha$ 1) is connected by a sharp  $\sim 90^\circ$  kink to a loop that adopts a partial, stretched-out helical conformation ( $\alpha$ 1'), amino acids 6–16) that occupies the groove formed by the JID and TAD to cap the central  $\beta$ -sheet (Fig. 1a and Extended Data Fig. 3a). In N-terminally truncated MYC3 (MYC3(44–238), which lacks  $\alpha$ 1' +  $\alpha$ 1), the JID rearranges to adopt a position similar to that of  $\alpha$ 1' in MYC3(5–242) to substitute for  $\alpha$ 1' to cap the  $\beta$ -sheet in the fold (Fig. 1b). We performed hydrogen deuterium exchange (HDX) experiments to detect the surface accessibility and structural dynamics of MYC3(5–242) in solution (Extended Data Fig. 4). While the central  $\beta$ -sheet has a very stable structure and is well protected from deuterium exchange, the  $\alpha$ 1/ $\alpha$ 1' helix region has a very high deuterium exchange rate, suggesting that it has a very dynamic structure and forms only transiently in solution. This is consistent with the high B-factor values of the  $\alpha$ 1/ $\alpha$ 1' helix in the MYC3(5–242) crystal structure (Extended Data Fig. 5). While peptides corresponding to the JID helix were not resolved in HDX experiments, the JID helix also has a high B-factor (Extended Data Fig. 5), indicating that its position is dynamic as well. The MYC3(5–242) and MYC3(44–238) apo crystal structures therefore probably represent structure snapshots of two or more alternative MYC3 conformations in solutions.

To crystallize a complex between MYC and the Jas motif, we synthesized a set of nine JAZ8, JAZ9 and JAZ12 Jas peptides of different lengths and complexed them with the above-mentioned set of the MYC N-terminal proteins. After extensive trials, we succeeded

<sup>1</sup>Laboratory of Structural Sciences and Laboratory of Structural Biology and Biochemistry, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA. <sup>2</sup>DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824, USA. <sup>3</sup>College of Plant Protection, Nanjing Agricultural University, No. 1 Weigang, 210095, Nanjing, Jiangsu Province, China. <sup>4</sup>Department of Biological Sciences, Western Michigan University, Kalamazoo, Michigan 49008, USA. <sup>5</sup>Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, USA. <sup>6</sup>Department of Molecular Therapeutics, Translational Research Institute, The Scripps Research Institute, Scripps Florida, Jupiter, Florida 33458, USA. <sup>7</sup>College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, Zhejiang, China. <sup>8</sup>Department of Molecular Pharmacology and Biological Chemistry, Life Sciences Collaborative Access Team, Synchrotron Research Center, Northwestern University, Argonne, Illinois 60439, USA. <sup>9</sup>Key Laboratory of Receptor Research, VARI-SIMM Center, Center for Structure and Function of Drug Targets, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. <sup>10</sup>Howard Hughes Medical Institute, Michigan State University, East Lansing, Michigan 48824, USA.

\*These authors contributed equally to this work.



**Figure 1 | Structures of MYC3 N terminus in apo and Jas peptide-bound states.** **a**, Apo MYC3(5–242). **b**, Apo MYC3(44–238). **c**, MYC3(5–242) bound to the 22-amino-acid Jas<sup>JAZ9</sup> motif peptide. Blue, JID; green, TAD; pink, Jas peptide. Dotted lines indicate the unresolved linkers that flank  $\alpha 1'/\alpha 1$  and the JID helices. **d**, MYC3(44–238) bound to the 22-amino-acid Jas<sup>JAZ9</sup> motif peptide. **e**, **f**, Overlay of the MYC3(44–238)–JAZ9 complex with apo MYC3(5–242) (**e**) and apo MYC3(44–238) (**f**). The complex structure is shown in colour overlaid on the apo structures in grey. Binding of the JAZ9 Jas22 helix (pink) displaces the  $\alpha 1'$  helix to become an integral part of the fold.

in obtaining crystals for MYC3(44–238) in complex with a 22-amino-acid Jas peptide (S218–M239) from JAZ9 (Jas22<sup>JAZ9</sup>). However, no crystals for MYC3(5–242)–Jas peptide complexes could be obtained, suggesting that Jas complexes with  $\alpha 1'/\alpha 1$  helix-containing MYC3 are less stable and/or conformationally more dynamic. To test this hypothesis, we generated a covalent fusion between MYC3(5–242) and Jas22<sup>JAZ9</sup> separated by a 12-amino-acid flexible linker to increase the stability of the MYC3(5–242) and Jas22<sup>JAZ9</sup> complex and reduce conformational flexibility. This fusion protein formed high-quality crystals and allowed us to solve the structure of the MYC3(5–242)–Jas22<sup>JAZ9</sup> complex (Fig. 1c) at a resolution of 2.4 Å (Extended Data Table 1). The most striking aspect of the MYC3(5–242)–Jas22<sup>JAZ9</sup> complex is that the Jas peptide formed a single, continuous helix that displaced the dynamic  $\alpha 1'/\alpha 1$  helix in apo MYC3(5–242) (Supplementary Video 1). Correspondingly, the JID helix rearranged its conformation and the displaced  $\alpha 1'/\alpha 1$  helix became almost completely disordered, suggesting the increase in disorder as the likely reason for the recalcitrance of this complex to crystallize. The Jas helix adopted a position in the groove that is almost superimposable with that of  $\alpha 1'$  (Fig. 1e). In this position the Jas helix is nestled between the  $\alpha 4$ -helix of the TAD and the strand and helix of the JID to make extensive interactions with both the JID and the TAD and to become an integral part of the structural fold. In addition, we

also determined the structure of MYC3(44–238), lacking the  $\alpha 1'/\alpha 1$  helix, in complex with Jas22<sup>JAZ9</sup> peptide at a resolution of 1.95 Å (Extended Data Table 1). In this structure, the Jas peptide also adopted a helical conformation and binds to the JID/TAD groove in the same way as seen in the MYC3(5–242)–Jas22<sup>JAZ9</sup> fusion complex (Fig. 1d, f). Both  $\alpha 1'/\alpha 1$  and the JID helix have long, unresolved linkers that appear to provide flexibility for their displacements/rearrangements by the Jas helix. Together, our analyses of the MYC3 apo and MYC3–JAZ complex structures indicate that occupancy of the groove (either by  $\alpha 1'$  or the JID helix in apo MYC3, or by the Jas helix in the MYC–JAZ complex) is critical for formation of the overall MYC N-terminal fold.

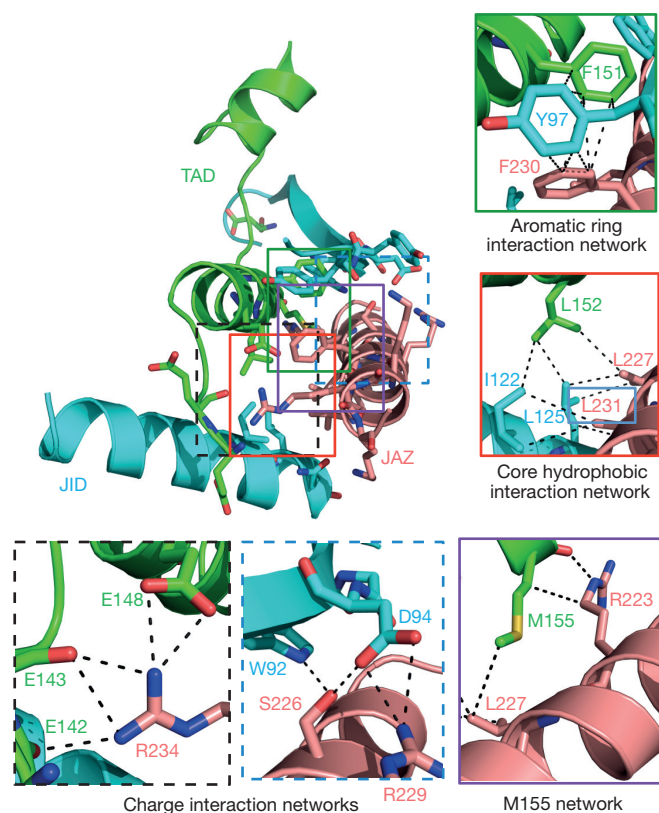
In the JAZ9–MYC3 complex, the JAZ peptide forms five main interaction networks with the TAD–JID surface (Fig. 2 and Extended Data Table 2): (1) R234 of JAZ9 forms salt bridges and hydrogen bonds with three glutamate residues (E142, E143 and E148) of the TAD; (2) R229 and S226 form charge and hydrogen bond interactions with JID D94 and W92; (3) L227 and L231 form a core hydrophobic network with TAD L152 and JID I122 and L125; (4) F230 interacts with JID Y97 and TAD F151 in an aromatic ring network; and (5) R223 and L227 have both hydrogen bond and/or hydrophobic interactions with TAD M155. In contrast, the first five amino acids (SVPQA) of the Jas motif that are critical for the co-receptor function of JAZ proteins (that is, hormone-dependent binding to COI1 (ref. 16)) made no critical interactions with MYC3, consistent with our yeast two-hybrid and AlphaScreen data (Extended Data Figs 1b, d and 2b). Consistent with the structural data, mutational analysis showed that key interface residues of Jas, JID and TAD have important roles in MYC–JAZ interactions (Fig. 3a, b and Extended Data Fig. 6). In addition, the structure also provides an explanation for the partial *in vivo* relief of MYC3 repression by the MYC3<sup>D94N</sup> mutation observed previously<sup>20</sup>, as MYC3<sup>D94N</sup> lost interaction with a subset of JAZs, including JAZ3, JAZ4 and JAZ9 (Extended Data Fig. 6).

Next, we transfected the MYC-responsive pJAZ2::GUS reporter<sup>21</sup> together with wild-type and mutant MYC3 expression plasmids into *Arabidopsis* protoplasts. As shown in Fig. 3c, mutant MYC3 proteins that were defective in interaction with multiple JAZ proteins (Extended Data Fig. 6) were partly relieved in repression (that is, increased reporter gene activity). Moreover, the extent at which mutations compromised MYC3 interactions with JAZ proteins correlated with the increase in reporter gene activity and the magnitude of changes in reporter gene activity could be further accentuated by expressing MYC3 mutant proteins from the strong cauliflower mosaic virus 35S promoter in *coi1-30* mutant protoplasts, in which all JAZ repressors are presumably stabilized (Fig. 3d). Together, these data validate the MYC3–JAZ9 complex structure and provide strong evidence that amino-acid interactions identified in the MYC3–JAZ9 complex structure are important for MYC3 repression *in planta*.

The Jas motif is required for its repressor function through interaction with MYC but also for its co-receptor function through interaction with COI1 (ref. 22). While the Jas<sup>JAZ9</sup> peptide in the MYC3 complex formed a continuous helix (Fig. 1c, d), representing the rest state of JAZ, the Jas<sup>JAZ1</sup> peptide in the previously determined COI1–jasmonate–Ile–Jas co-receptor structure (Protein Data Bank accession number 3OGL) adopted a bipartite conformation with an N-terminal part stretched to form a distinct loop region followed by a shorter C-terminal helix<sup>16</sup>, as illustrated by the structural alignment in Fig. 4a, b and Supplementary Video 2. In addition to the Jas<sup>JAZ9</sup>–MYC3 complex, we solved the structure of the Jas<sup>JAZ1</sup>–MYC3 complex. As shown in the structure alignment in Extended Data Fig. 7, the Jas helices of JAZ9 and JAZ1 overlap very well, confirming that the Jas conformational change between MYC-bound (resting stage) and COI1-bound (hormone-activated stage) is probably common in MYC interaction with different JAZ transcriptional repressors.

In the Jas<sup>JAZ1</sup>–COI1 complex, the loop region of the Jas<sup>JAZ1</sup> helix is formed by the five moderately conserved N-terminal amino

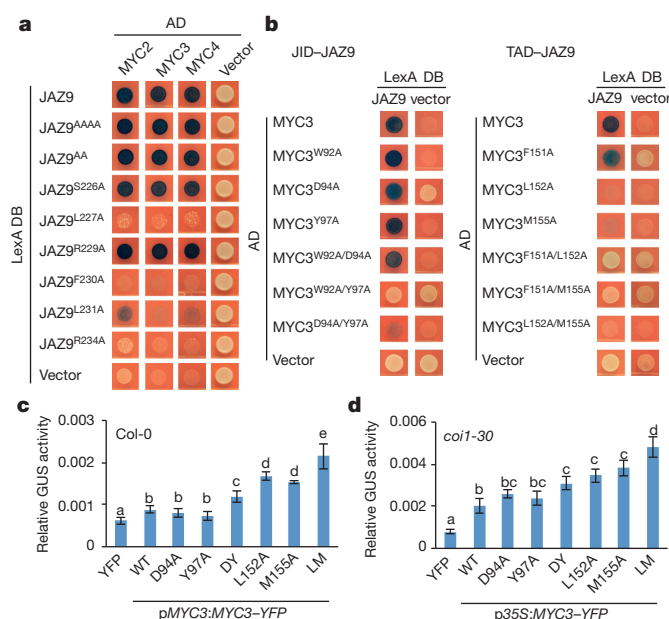




**Figure 2 | Jas peptide forms extensive interactions with the JID-TAD surface in MYC3.** JID-TAD-Jas structure with important interacting residues shown in stick presentation. Details of key interaction networks are shown as solid boxes (hydrophobic interaction networks) or dashed boxes (charge interaction networks). For clarity, not all interacting residues are shown in the detail boxes, compared with the non-boxed overview figure.

acids of the Jas motif (Extended Data Fig. 1c) that directly interact with the jasmonate-Ile hormone (Fig. 4a, b) and is required for Jas-jasmonate-Ile-CO1 co-receptor complex formation<sup>16</sup>. When we mutated the corresponding N-terminal amino acids of JAZ9 to alanine (JAZ9-4A and JAZ9-AA; Fig. 4c), JAZ9 lost interaction with CO11 in yeast two-hybrid assays, but not with MYC3 (Fig. 4d), consistent with the MYC3 complex structure. Mutations in the middle of the Jas motif (S226A-R234A) affected binding to both MYC3 and CO11, albeit to different degrees. In addition, residues that are C-terminal to the Jas motif enhance JAZ9 interaction with CO11 in yeast two-hybrid assays, but are not critical for its interaction with MYC3 (Fig. 4e, f), which is consistent with a previous study of JAZ2, JAZ3 and JAZ10 interactions with CO11 and MYC2 (ref. 23). Together, these results indicate that CO11 and MYC3 potentially compete for binding to the central part of the Jas motif, but that CO11 makes additional critical interactions with JAZ9 outside the MYC3-interacting region, including the previously unrecognized hormone-dependent unwinding of the N-terminal helix of the Jas motif (Supplementary Video 2). These additional interactions may allow CO11 to drive JAZ ubiquitination and dissociation of the extensive JAZ-MYC3 interaction upon jasmonate-Ile stimulation.

MED25 is a subunit of the Mediator complex that recruits RNA polymerase II to the promoters of jasmonate-responsive genes<sup>24</sup> and is required for various jasmonate responses<sup>13,15,25</sup>, including *Arabidopsis* susceptibility to *Pseudomonas syringae* bacterial infection (Extended Data Fig. 7c) and jasmonate-induced inhibition of *Arabidopsis* root growth (Extended Data Fig. 7d). We found that MYC3(44-238) also directly binds MED25 and that a fragment (amino acids 540-680) encompassing the MED25 activator interaction domain (ACID) is sufficient to bind to MYC3 (Fig. 5a), analogous to what has previously been reported for MYC2 (ref. 13). Since the MYC3 TAD makes critical

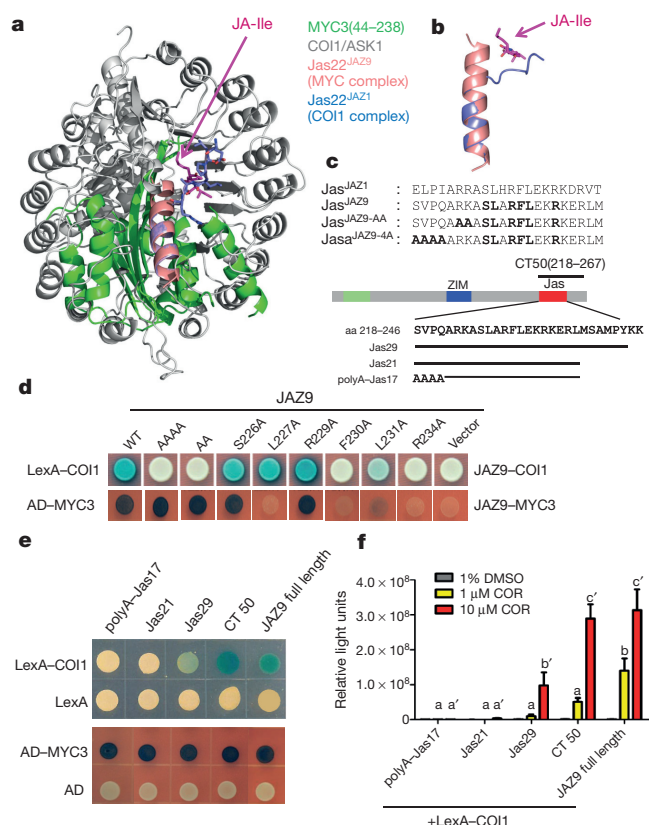


**Figure 3 | Mutational analysis of the JAZ9-MYC3 interaction.** **a**, Yeast two-hybrid analysis of the interaction between JAZ9 mutant proteins and wild-type MYC2, MYC3 and MYC4 proteins. JAZ9<sup>AAAA</sup> contains S218A, V219A, P220A and Q221A mutations, and JAZ9<sup>AA</sup> contains R223A and K224A mutations. The experiment was repeated three times with same results. DB, Y2H bait vector carrying LexA DNA binding domain. AD, Y2H prey vector carrying the B42 activation domain. **b**, Yeast two-hybrid analysis between wild-type JAZ9 protein and MYC3 proteins with mutations in the JID (left) or TAD (right). The experiment was repeated three times with same results. **c**, **d**, Alanine replacements of JAZ9-interacting amino acids of MYC3 increase MYC3 target gene expression in wild-type Col-0 and *col1-30* mutant plants, respectively. *Arabidopsis* protoplasts from wild type (c) or *col1-30* mutants (d) were transfected with an MYC3-responsive JAZ2::GUS reporter together with yellow fluorescent protein (YFP) alone or MYC3-YFP constructs under control of the native MYC3 (c) or cauliflower mosaic virus 35S (d) promoter as indicated. A 35S::LUC reporter construct was co-transfected as a control. GUS activities were normalized to the luciferase activity. Data shown are means of four independent transfections ( $n = 4$  biological replicates; error bars, s.d.). Different letters above the columns indicate the significant differences from each other ( $P < 0.05$ ) in pJAZ2::Gus reporter activities from transient expression of indicated MYC3 variants, as determined by Tukey-Kramer multiple comparison analysis. DY, the D94A/Y97A double mutant; LM, the L152A/M155A double mutant. The experiment was repeated three times with similar results.

interactions with JAZ repressors and is required for MYC3-JAZ9 complex formation (Figs 2 and 3b and Extended Data Fig. 6), we explored the intriguing possibility that MYC3 binding of JAZ9 and MED25 is mutually exclusive. To test this prediction in a defined system, we performed AlphaScreen interaction assays between MED25(407-680) and both MYC3(44-238) and MYC3(5-242) in the presence of increasing amounts of untagged Jas22<sup>JAZ9</sup> peptide. As shown in Fig. 5b, the JAZ peptide competitively inhibited the MYC-MED25 interaction with an  $IC_{50}$  of ~420 nM (MYC3(44-238)) and ~490 nM (MYC3(5-242)). We further tested competition *in planta* by transiently expressing combinations of tagged MED25, JAZ9 and MYC3 in *Nicotiana tabacum* leaves. As shown in Fig. 5c, co-immunoprecipitation of MED25 with MYC3 was strongly reduced upon co-expression of JAZ9. Together, these results demonstrate that the Jas motif of JAZ proteins and the ACID domain of MED25 probably bind to a shared MYC3 surface, and that JAZ repressors can compete with MYC3 for interaction with MED25 (and possibly other co-activators) *in vitro* and *in planta*.

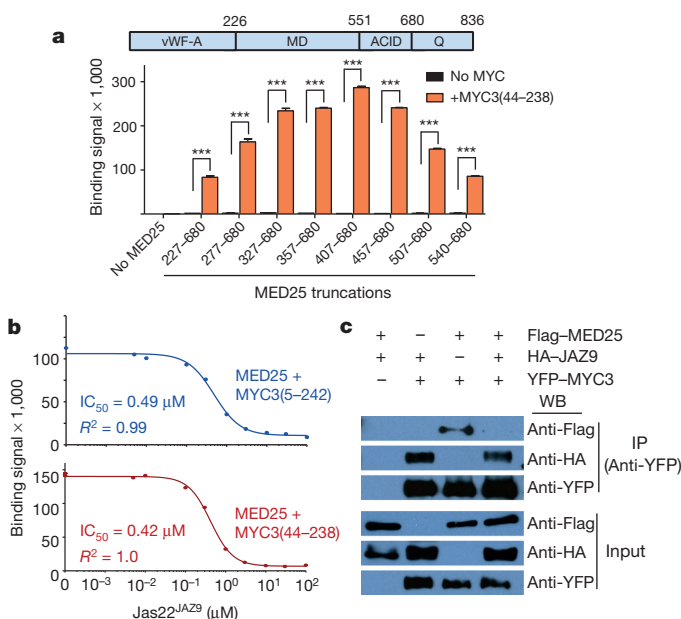
In the past decade, despite the identification of analogous hormone perception and transcriptional gene regulation that underpins several hormone signal transduction pathways in plants<sup>26</sup>, no crystal structures of the transcriptional-repressor-transcription-factor complexes





**Figure 4 | Distinct conformations of the Jas helix in the COI1-JAZ co-receptor complex versus the JAZ-MYC complex.** **a**, Structure overlay of the MYC3-Jas<sup>JAZ9</sup> complex with the COI1/ASK-jasmonate-Ile-Jas<sup>JAZ1</sup> complex (Protein Data Bank accession number 3OGL). **b**, Close up of the Jas motif overlay from the MYC3-Jas<sup>JAZ9</sup> complex (pink) and from the COI1/ASK-jasmonate-Ile-Jas<sup>JAZ1</sup> complex (blue). **c**, Alignment of the Jas motifs from JAZ1 and JAZ9. Mutationally analysed conserved single residues and N-terminal amino-acid stretches in JAZ9 are highlighted in bold. The position and sequence of JAZ9 fragments are indicated below the simplified JAZ9 diagram. **d**, Yeast two-hybrid analysis of the interaction between JAZ9 and either COI1 (top) or MYC3 (bottom). The experiment was repeated three times with the same results. **e**, Yeast two-hybrid analysis of the interaction between JAZ9 C-terminal fragments and either COI1 (top) or MYC3 (bottom). One micromolar coronatine (COR) was used in yeast two-hybrid assays for the COI1-JAZ interaction. The experiment was repeated three times with similar results. **f**, Quantitative yeast two-hybrid analysis of the interaction between JAZ9 truncations and COI1, with  $\beta$ -galactosidase reporter gene activity determined by Beta-Glo assay ( $n = 3$  biological replicates; error bars, s.d.). Different letters above the columns indicate significant differences from each other ( $P < 0.05$ ) in COI1 interaction with indicated JAZ9 fragments at a given concentration of COR (that is, 1  $\mu$ M or 10  $\mu$ M), as determined by two-way analysis of variance (ANOVA) with Bonferroni post-test. The experiment was repeated three times with similar results.

have been solved. The crystal structure of the MYC-JAZ complex reported here therefore provides the first structural insight into the mechanism of transcriptional repression in plant hormone signalling. Our structural, biochemical and *in planta* analyses suggest that JAZ repressors use a novel dual repression mechanism, which involves not only epigenetic modifications of the target gene chromatin structure through TOPLESS co-repressors, as demonstrated previously<sup>27</sup>, but also direct inhibition of MYC binding to MED25 (and possibly other co-activators), as an integral part of a mechanism of preventing transcriptional activation of jasmonate response genes (Extended Data Fig. 8). In addition, we have discovered distinct JAZ conformations in the MYC-JAZ resting complex versus the JAZ-COI1 hormone-activated complex<sup>16</sup>, providing the first structural insight into the switch mechanism between transcriptional



**Figure 5 | Jas motif peptide competitively inhibits the MYC3-MED25 interaction.** **a**, MYC3 interacts with the ACID domain of MED25. Note that sequences N-terminal to ACID contribute to binding and/or stability. AlphaScreen interaction assay between biotinylated MYC3(44–238) and His6Sumo-MED25 fragments ( $n = 3$  technical replicates; error bars, s.d.). \*\*\*Significant differences ( $P < 0.001$ ) compared with the no-MYC control by Student's *t*-test. vWF-A, von Willebrand factor A domain (responsible for mediator binding); MD, middle domain; Q, Q-rich C-terminal region. The experiment was repeated three times with similar results. **b**, Jas<sup>JAZ9</sup> peptide competes with MED25 for MYC3 binding. AlphaScreen competition assay ( $n = 3$  technical replicates; error bars, s.d.). Jas22<sup>JAZ9</sup>: untagged JAZ9(218–239). MED25: His6Sumo-MED25(407–680). MYC3(5–242): biotin-MYC3(5–242). MYC3(44–238): biotin-MYC3(44–238). The experiment was repeated three times with similar results. An enlarged version of **b** with visible error bars is shown in Supplementary Fig. 1, and associated original data in Supplementary Table 1. **c**, Interference of the MYC3-MED25 interaction by JAZ9 *in planta*. Flag-MED25 (+), haemagglutinin epitope tag (HA)-JAZ9 (+) and YFP-MYC3 (+), and respective vector controls carrying Flag, HA or YFP tags (–) under control of the cauliflower mosaic virus 35S promoter were transiently expressed in *N. tabacum* leaves. Protein extracts were immunoprecipitated (IP) with an anti-YFP antibody and analysed by western blot (WB) with HA-, Flag-, or YFP-specific antibodies. The experiment was repeated three times with similar results. The original blots from which the images were cropped are shown in Supplementary Fig. 2.

repression and hormone-dependent transcriptional activation in a major plant hormone signalling pathway.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 November 2014; accepted 15 June 2015.

Published online 10 August; corrected online 9 September 2015 (see full-text HTML version for details).

- Browse, J. Jasmonate: preventing the maize tassel from getting in touch with his feminine side. *Sci. Signal.* **2**, pe9 (2009).
- Wager, A. & Browse, J. Social network: JAZ protein interactions expand our knowledge of jasmonate signaling. *Front. Plant Sci.* **3**, 41 (2012).
- Chini, A. *et al.* The JAZ family of repressors is the missing link in jasmonate signalling. *Nature* **448**, 666–671 (2007).
- Farmer, E. E., Almeras, E. & Krishnamurthy, V. Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Curr. Opin. Plant Biol.* **6**, 372–378 (2003).
- Liechti, R. & Farmer, E. E. The jasmonate pathway. *Science* **296**, 1649–1650 (2002).
- Pauwels, L., Inze, D. & Goossens, A. Jasmonate-inducible gene: what does it mean? *Trends Plant Sci.* **14**, 87–91 (2009).
- Yan, J. *et al.* The *Arabidopsis* CORONATINE INSENSITIVE1 protein is a jasmonate receptor. *Plant Cell* **21**, 2220–2236 (2009).

8. Fonseca, S. *et al.* (+)-7-*iso*-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nature Chem. Biol.* **5**, 344–350 (2009).
9. Katsir, L., Schilmiller, A. L., Staswick, P. E., He, S. Y. & Howe, G. A. COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc. Natl Acad. Sci. USA* **105**, 7100–7105 (2008).
10. Thines, B. *et al.* JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. *Nature* **448**, 661–665 (2007).
11. Xie, D.-X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280**, 1091–1094 (1998).
12. Yan, Y. *et al.* A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* **19**, 2470–2483 (2007).
13. Chen, R. *et al.* The *Arabidopsis* Mediator subunit MED25 differentially regulates jasmonate and abscisic acid signaling through interacting with the MYC2 and ABI5 transcription factors. *Plant Cell* **24**, 2898–2916 (2012).
14. Fernandez-Calvo, P. *et al.* The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* **23**, 701–715 (2011).
15. Cevik, V. *et al.* MEDIATOR25 acts as an integrative hub for the regulation of jasmonate-responsive gene expression in *Arabidopsis*. *Plant Physiol.* **160**, 541–555 (2012).
16. Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1–JAZ co-receptor. *Nature* **468**, 400–405 (2010).
17. Melcher, K. The strength of acidic activation domains correlates with their affinity for both transcriptional and non-transcriptional proteins. *J. Mol. Biol.* **301**, 1097–1112 (2000).
18. Sun, X., Rikkerink, E. H., Jones, W. T. & Uversky, V. N. Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* **25**, 38–55 (2013).
19. Triezenberg, S. J. Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.* **5**, 190–196 (1995).
20. Smolen, G. A., Pawlowski, L., Wilensky, S. E. & Bender, J. Dominant alleles of the basic helix-loop-helix transcription factor ATR2 activate stress-responsive genes in *Arabidopsis*. *Genetics* **161**, 1235–1246 (2002).
21. Figueroa, P. & Browse, J. The *Arabidopsis* JAZ2 promoter contains a G-Box and thymidine-rich module that are necessary and sufficient for jasmonate-dependent activation by MYC transcription factors and repression by JAZ proteins. *Plant Cell Physiol.* **53**, 330–343 (2012).
22. Melotto, M. *et al.* A critical role of two positively charged amino acids in the Jas motif of *Arabidopsis* JAZ proteins in mediating coronatine- and jasmonoyl isoleucine-dependent interactions with the COI1 F-box protein. *Plant J.* **55**, 979–988 (2008).
23. Chung, H. S. *et al.* Alternative splicing expands the repertoire of dominant JAZ repressors of jasmonate signaling. *Plant J.* **63**, 613–622 (2010).
24. Kazan, K. & Manners, J. M. MYC2: the master in action. *Mol. Plant* **6**, 686–703 (2013).
25. Kidd, B. N. *et al.* The Mediator complex subunit PFT1 is a key regulator of jasmonate-dependent defense in *Arabidopsis*. *Plant Cell* **21**, 2237–2252 (2009).
26. Santner, A. & Estelle, M. Recent advances and emerging trends in plant hormone signalling. *Nature* **459**, 1071–1078 (2009).
27. Pauwels, L. *et al.* NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* **464**, 788–791 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This research is supported by the Gordon and Betty Moore Foundation (GBMF3037, to S.Y.H.), the China Scholarship Council (to F.Z.), Van Andel Research Institute (to H.E.X. and K.M.), the National Institutes of Health (R01 GM102545 to K.M. and R01AI060761 to S.Y.H.), and the Department of Energy (the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science Grant DE-FG02-91ER20021 (to S.Y.H.)). We thank S. Grant for administrative support and staff members of the Life Science Collaborative Access Team of the Advanced Photon Source for assistance in data collection at the beam lines of sector 21, which is in part funded by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). Use of the Advanced Photon Source was supported by the Office of Science of the US Department of Energy, under contract number DE-AC02-06CH11357. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We also thank G. Howe and K. Aung for reading the manuscript.

**Author Contributions** H.E.X., K.M., M.Z. and S.Y.H. conceived the project and designed experiments. F.Z., J.Y., J.K., L.Z., V.Q.L., X.F.X., X.E.Z., J.C., J.B. and P.R.G. performed and/or interpreted experiments. K.M., F.Z., J.Y. and S.Y.H. wrote the manuscript with support from all authors.

**Author Information** Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession numbers 4RRU, 4RQW, 4RS9, 4YZ6 and 4YWC (see Extended Data Table 1). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.E.X. (Eric.Xu@vai.org), K.M. (Karsten.Melcher@vai.org) or S.Y.H. (hes@msu.edu).

## METHODS

**Protein preparation.** Wild-type MYC3(44–238) was expressed as a fusion protein with a cleavable N-terminal His6Sumo tag from a modified pSUMO (LifeSensors) expression vector. BL21 (DE3) cells transformed with the expression plasmid were grown in Luria–Bertani broth at 16 °C to an absorbance ( $A_{600\text{ nm}}$ ) of  $\sim 1.0$  and induced with 0.1 mM IPTG for 16 h. Cells were harvested, resuspended in 100 ml extract buffer (20 mM Tris, pH 8.0, 200 mM NaCl, and 10% glycerol) per 6 l of cells and passed three times through a French press with pressure set at 1,000 Pa. The lysate was centrifuged at 40,900g in a Sorvall SS34 rotor for 30 min, and the supernatant was loaded on a 50 ml nickel HP column. The column was washed with 600 ml of 10% buffer B (20 mM Tris, pH 8.0, 200 mM NaCl, 500 mM imidazole and 10% glycerol) and eluted with 200 ml of 50% buffer B, followed by 100 ml of 100% buffer B. The eluted His6Sumo–MYC3(44–238) was dialysed against extract buffer and cleaved overnight with SUMO protease at a protease:protein ratio of 1:1,000 at 4 °C. The cleaved His6Sumo tag was removed by passing through a 5 ml nickel HP column, and the protein was further purified by chromatography through a HiLoad 26/60 Superdex 200 gel filtration column in 25 mM Tris, pH 8.0, 200 mM ammonium acetate, 1 mM dithiothreitol and 1 mM EDTA. To prepare the protein–ligand complex, we mixed Jas22<sup>JAZ1</sup>, Jas22<sup>JAZ8</sup>, Jas22<sup>JAZ9</sup> or Jas22<sup>JAZ12</sup> peptides with purified MYC3(44–238) proteins at a 1.5:1 molar ratio.

The expression and purification of MYC3(5–242) followed the same method as for MYC3(44–238) described above. To prepare the protein–ligand complex, we mixed Jas22<sup>JAZ1</sup>, Jas22<sup>JAZ8</sup>, Jas22<sup>JAZ9</sup> or Jas22<sup>JAZ12</sup> peptides with purified MYC3(5–242) proteins at a 1.5:1 molar ratio. To prepare MYC3(44–238) selenomethionyl (Se-Met) protein for phase determination, we followed the same methods as described previously<sup>28</sup>. Purification of Se-Met MYC3(44–238) proteins followed the same protocol as for MYC3(44–238) native protein except that the procedure was performed more quickly to avoid protein oxidation. The MYC3(5–242)–Jas22<sup>JAZ9</sup> complex was constructed as a fusion protein with His6Sumo–MYC3(5–242) at the N terminus and Jas22<sup>JAZ9</sup> at the C terminus, separated by a flexible GSAGSAGSAGSA (4×GSA) linker (His6Sumo–MYC3(5–242)–4×GSA–Jas22<sup>JAZ9</sup>). The expression and purification of the fusion protein followed the same methods as for MYC3(44–238).

Small-scale purification of His6Sumo-tagged MYC2/3/4 protein fragments (including JID and TAD domain) for binding studies with Jas peptides followed the same methods as for MYC3(44–238), except that the His6Sumo tag was not removed. Small-scale purification of His6Sumo-tagged MED25 protein fragments (including the ACID domain) for binding studies with biotinylated MYC3(44–238) followed the same methods as for MYC3(44–238), except that the His6Sumo tag was not removed. To express and purify biotinylated MYC3(44–238) and MYC3(5–242) protein for binding studies (Fig. 5a) and JAZ competition assays (Fig. 5b), we followed the methods described previously<sup>29</sup>.

**Crystallization.** The apo-MYC3(5–242) crystals were grown at 20 °C in sitting drops containing 0.2  $\mu$ l of purified MYC3(5–242) protein at a concentration of 10 mg ml<sup>−1</sup> and 0.2  $\mu$ l of well solution containing 0.2 M magnesium chloride, 0.1 M Tris, pH 8.5, 30% (w/v) polyethylene glycol 4000 for 3 days. The Se-Met MYC3(44–238) crystals were grown at 20 °C in sitting drops containing 0.2  $\mu$ l of the purified protein at a concentration of 15 mg ml<sup>−1</sup> and 0.2  $\mu$ l of well solution containing 0.2 M sodium chloride, 0.1 M Bis-Tris, pH 5.5 and 25% (w/v) polyethylene glycol 3,350. Crystals of about 100  $\mu$ m in length appeared in 3 days. The MYC3(44–238)–Jas22<sup>JAZ9</sup> complex crystals were grown at 20 °C in sitting drops containing 0.2  $\mu$ l of the purified complex proteins at a concentration of 15 mg ml<sup>−1</sup> and 0.2  $\mu$ l of well solution containing 0.2 M magnesium chloride, 0.1 M Tris, pH 8.5 and 30% (w/v) polyethylene glycol 4000 for 3 days. The MYC3(44–238)–Jas22<sup>JAZ1</sup> complex crystals were grown at 20 °C in sitting drops containing 0.2  $\mu$ l of the purified complex proteins at a concentration of 15 mg ml<sup>−1</sup> and 0.2  $\mu$ l of well solution containing 3.5 M sodium formate. Crystals of about 80  $\mu$ m in length appeared in 2 days. The MYC3(5–242)–Jas22<sup>JAZ9</sup> fusion protein crystals were grown at 20 °C in sitting drops containing 0.2  $\mu$ l of the purified fusion proteins at a concentration of 15 mg ml<sup>−1</sup> and 0.2  $\mu$ l of well solution containing 0.2 M magnesium nitrate, 20% (w/v) polyethylene glycol 3350. Crystals of about 100  $\mu$ m in length appeared in 3 days. All crystals were serially transferred to the well solution with 20% (v/v) ethylene glycol before flash freezing in liquid nitrogen.

**Data collection and structure determination.** Data collections were performed at sector 21-ID (LS-CAT) beam lines of the Advanced Photon Source synchrotron using single native MYC3(5–242) crystals, MYC3(44–238)–Jas22<sup>JAZ9</sup> complex crystals and Se-Met-substituted MYC3(44–238) crystals at different wavelengths. All diffraction data were processed using XDS<sup>30</sup>, and merged using Aimless of the CCP4 suite<sup>31</sup>. Initial phasing was tried by using the SAD method based on anomalous diffraction of sulfur atoms as previously described<sup>32</sup>. S-SAD phasing using a combined data set of 11 native MYC3(5–242) crystals collected at 1.77 Å was not successful, probably because of the non-isomorphism of the individual crystals. To

solve the phase problem, the Se-Met-substituted MYC3(44–238) crystals were prepared and a single data set was collected at a peak wavelength of 0.9787 Å (Extended Data Table 1). Se-SAD phasing was performed by using the Phenix Autosol program. Five out of six selenium atoms were found with a figure of merit (FOM) value of 0.41. The Phenix autobuild program generated an initial model of 286 residues with a value of  $R_{\text{work}}/R_{\text{free}}$  of 0.35/0.40. Further model improvements were performed using Coot<sup>33</sup> and refined using the Refmac program in CCP4 (ref. 34) to a final model with an  $R$  factor of 0.21 and an  $R_{\text{free}}$  factor of 0.26. The MYC3(5–242) apo structure, MYC3(5–242)–Jas22<sup>JAZ9</sup>, MYC3(44–238)–Jas22<sup>JAZ9</sup> and MYC3(44–238)–Jas22<sup>JAZ1</sup> complex structures were solved by using the molecular replacement program Phaser<sup>35</sup> with the Se-Met MYC3(44–238) structure as a search model. The model for the  $\alpha 1'/\alpha 1$  helices for the MYC3(5–242) structure, for the Jas22<sup>JAZ9</sup> peptide for the MYC3(44–238)–Jas22<sup>JAZ9</sup> and MYC3(5–242)–Jas22<sup>JAZ9</sup> complex structures and for the Jas22<sup>JAZ1</sup> peptide for the MYC3(44–238)–Jas22<sup>JAZ1</sup> complex structure were built on the basis of the electron density maps using Coot.

**AlphaScreen luminescence proximity assays.** *In vitro* interactions between MYC3 and Jas peptides or MED25 fragments were assessed by luminescence proximity AlphaScreen (PerkinElmer) technology as described previously<sup>29,36,37</sup>. Reactions contained 50 nM His6Sumo–MYC3 protein bound to nickel-acceptor beads and 50 nM synthesized biotinylated Jas peptides bound to streptavidin donor beads (Extended Data Fig. 1d and 2b) or 50 nM His6Sumo–MED25–ACID protein bound to nickel-acceptor beads and 50 nM biotin–MYC3(44–238) bound to streptavidin donor beads (Fig. 5a). The results were based on an average of three experiments with standard errors typically less than 10% of the measurement.

For the competition assay (Fig. 5b), non-biotinylated Jas22<sup>JAZ9</sup> peptide was added into the reaction at concentrations of 0, 5, 10, 100, 300, 1,000, 3,000, 10,000, 30,000 and 100,000 nM. The results were based on an average of three experiments with standard errors typically less than 10% of the measurement.

**Mutagenesis.** Site-directed mutagenesis was performed using the QuickChange Method (Agilent). Mutations for all plasmid constructs were confirmed by sequencing.

**Yeast two-hybrid assays.** Most of the constructs for yeast two-hybrid assays used in this study were described previously<sup>22,38,39</sup>. The coding sequences of full-length MYC2, MYC3 and MYC4, and MYC N-terminal and C-terminal fragments were PCR-amplified and cloned into pGilda (Clontech). PCR-based deletions were conducted following the manufacturer's protocol (Stratagene). Detailed protocols for yeast two-hybrid assays were described previously<sup>38</sup>.

**Plant materials and growth conditions.** *Arabidopsis* plants used in this study were previously described<sup>39</sup> or were ordered from the *Arabidopsis* Biological Resource Center ([www.arabidopsis.org](http://www.arabidopsis.org)). *Arabidopsis* seeds were stratified for 3 days at 4 °C before planting. The soil-grown plants were placed in a controlled growth chamber at 23 °C with a 12-h-day (80  $\mu$ mol s<sup>−1</sup> m<sup>−2</sup> cool-white fluorescent light)/12-h-night cycle.

**Transient expression in tobacco leaves and *Arabidopsis* protoplasts.** For transient expression in tobacco leaves, coding sequences of JAZ9, MED25 and MYC3 were cloned into pJYP003, pJYP011 and pJYP018 (J.Y. and S.Y.H., unpublished observations), respectively, to create p35S:3×HA–JAZ9, p35S:3×Flag–MED25 and p35S:YFP–MYC3 fusion constructs, which were transfected as previously described<sup>39</sup>. Protein extracts were immunoprecipitated with an anti-YFP antibody and analysed by western blot with anti-HA, -Flag or -YFP antibodies as previously described<sup>39</sup>. For transient expression in *Arabidopsis* mesophyll protoplasts, MYC3 (no stop codon) with or without its promoter (the 2-kb sequence upstream of the start codon) was PCR-amplified and cloned into pENTR-D/TOPO vector (Invitrogen) to create entry clones. Then, the MYC3 or pMYC3:MYC3 inserts were introduced into pSAT4A-DEST-Venus or pBR-DEST-Venus (J.Y. and S.Y.H., unpublished observations) to create the p35S:MYC3–YFP or pMYC3–MYC3–YFP constructs. The JAZ2 promoter (the 2-kb sequence upstream of its start codon) was cloned into pBR-Gus (J.Y. and S.Y.H., unpublished observations) to create pJAZ2:GUS reporter constructs. The transient expression assays using pBS-35S-Luc as transfection control followed a published protocol<sup>40</sup>. A 35S:LUC reporter construct was co-transfected as a control. GUS activities were normalized to the luciferase activity.

**Root growth inhibition assay.** *Arabidopsis* wild-type (Col-0) and *med25* (*pft1-2*; SALK\_129555)<sup>25</sup> seedlings were used for the root growth inhibition assay. Seeds were surface-sterilized, stratified at 4 °C and germinated on half-strength Murashige and Skoog agar plates containing 1  $\mu$ M, 3  $\mu$ M or 10  $\mu$ M MeJA or 0.1% DMSO (control). Plants were placed vertically in a growth chamber (16 h light/8 h dark light cycle, 100  $\mu$ E s<sup>−1</sup> m<sup>−2</sup> light intensity) for 10 days before pictures were taken, and root lengths were measured with ImageJ software (<http://rsbweb.nih.gov/ij/>).

**Bacterial infection.** *Pseudomonas syringae* pathovar *tomato* (*Pst*) DC3000 infection assays in *Arabidopsis* plants were performed as described previously<sup>41</sup>. Briefly, 5-week-old *Arabidopsis* plants were dip-inoculated with *Pst* DC3000 bacterial



suspension ( $1 \times 10^8$  colony-forming units per millilitre, 0.025% Silwet-L77). Bacterial growth was determined 3 days after inoculation.

**HDX mass spectrometry.** HDX of MYC3(5–242) in 20 mM Tris pH 8.0, 200 mM ammonium acetate, 1 mM EDTA and 7% glycerol was performed at 4 °C using an automated system described previously<sup>42</sup>. Briefly, protein was incubated in a D<sub>2</sub>O buffer for a range of exchange times from 10 s to 1 h before quenching the deuterium exchange reaction with an acidic quench solution (pH 2.4) containing 3 M urea and 1% TFA. All mixing and digestions were performed on a LEAP Technologies Twin HTS PAL liquid handling robot housed inside a temperature-controlled fridge. Protein digestion was performed in-line with chromatography using an immobilized pepsin column. Mass spectra were acquired on a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (ThermoFisher Scientific). Percentage deuterium exchange values for peptide isotopic envelopes at each time point were calculated and processed using Workbench software<sup>43</sup>.

The unit of measurement represented as a single value is the percentage deuterium incorporation, which is determined by initially calculating the intensity weighted average (centroid) of all spectral data within defined *m/z* limits. The percentage deuterium incorporation is then determined by comparing the result to defined minimum (0%) and maximum (100%) *m/z* values for each peptide. The minimum and maximum *m/z* values are determined using experimentally observed undeuterated and fully deuterated controls.

The data representing each peptide are reduced to single values in the following manner. For each sample, the three individual time-point replicates of the percentage deuterium incorporation (done in triplicate) at each time point are averaged. The mean of these values is then presented as a single value, representing the overall change in deuterium incorporation for the sample. The first number in brackets is the representation of the propagation of error for the sample, which is determined by a root mean squared approach using the standard deviations from each individual time point. The second number in brackets is the charge state of the detected peptide.

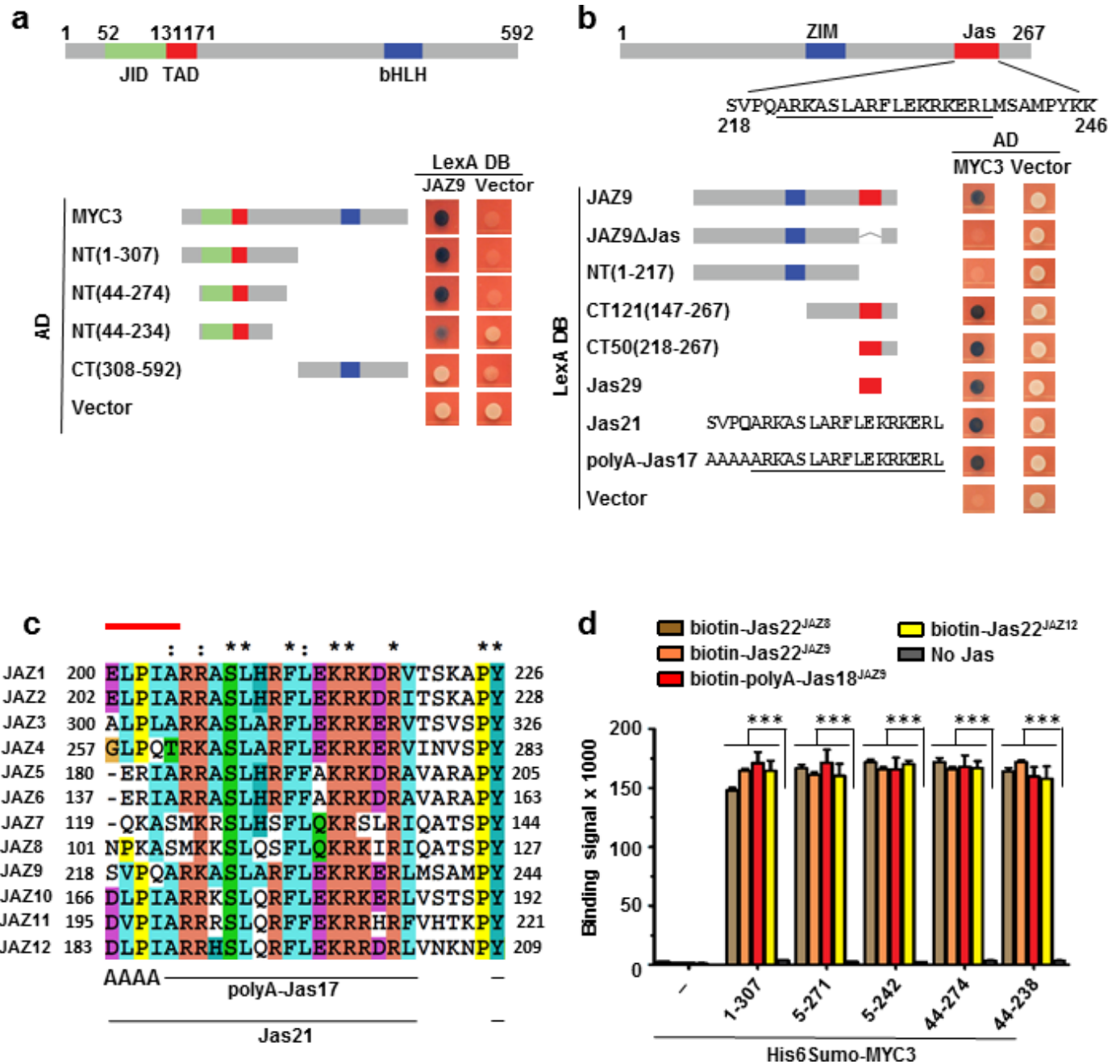
Extended Data Fig. 4a shows cumulative peptides fragments that were detected in tandem mass spectrometry. Shorter fragments (four to ten residues) provide higher resolution information than longer peptides. Therefore, they supersede longer fragments (more than ten residues) and were used to manually overlay onto the atomic structure as in the case of Fig. 2b. No subtraction was used. The peptide set used for structural overlay contained the shortest fragments from the complete data set (all peptides).

**Data analysis, statistics and experimental repeats.** The specific statistical method used, the sample size and the results of statistical analyses are described

in the relevant figure legends. No statistical methods were used to predetermine sample size. Sample size was determined on the basis of experimental trials and in consideration of previous publications on similar experiments to allow for confident statistical analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. All reported results were reproduced in at least three independent experiments.

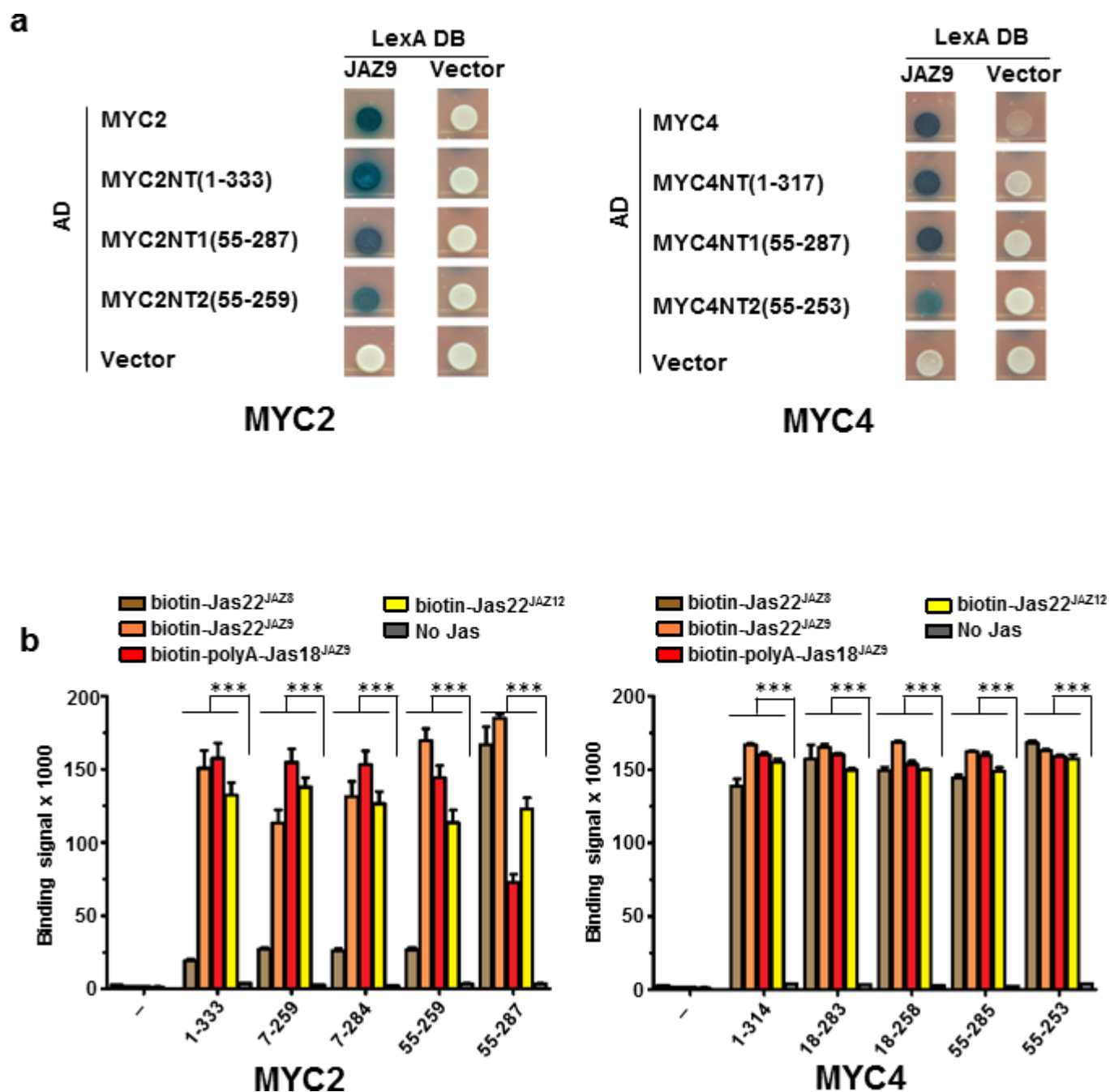
28. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
29. Melcher, K. *et al.* A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* **462**, 602–608 (2009).
30. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
31. Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
32. Ke, J. *et al.* Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nature Struct. Mol. Biol.* **20**, 1377–1382 (2013).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
36. Suino, K. *et al.* The nuclear xenobiotic receptor CAR; structural determinants of constitutive activation and heterodimerization. *Mol. Cell* **16**, 893–905 (2004).
37. Xu, H. E. *et al.* Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPAR $\alpha$ . *Nature* **415**, 813–817 (2002).
38. Withers, J. *et al.* Transcription factor-dependent nuclear localization of a transcriptional repressor in jasmonate hormone signaling. *Proc. Natl Acad. Sci. USA* **109**, 20148–20153 (2012).
39. Yang, D. L. *et al.* Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. *Proc. Natl Acad. Sci. USA* **109**, E1192–E1200 (2012).
40. Yoo, S. D., Cho, Y. H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2**, 1565–1572 (2007).
41. Yao, J., Withers, J. & He, S. Y. *Pseudomonas syringae* infection assays in *Arabidopsis*. *Methods Mol. Biol.* **1011**, 63–81 (2013).
42. Chalmers, M. J. *et al.* Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **78**, 1005–1014 (2006).
43. Pascal, B. D. *et al.* HDX workbench: software for the analysis of H/D exchange MS data. *J. Am. Soc. Mass Spectrom.* **23**, 1512–1521 (2012).





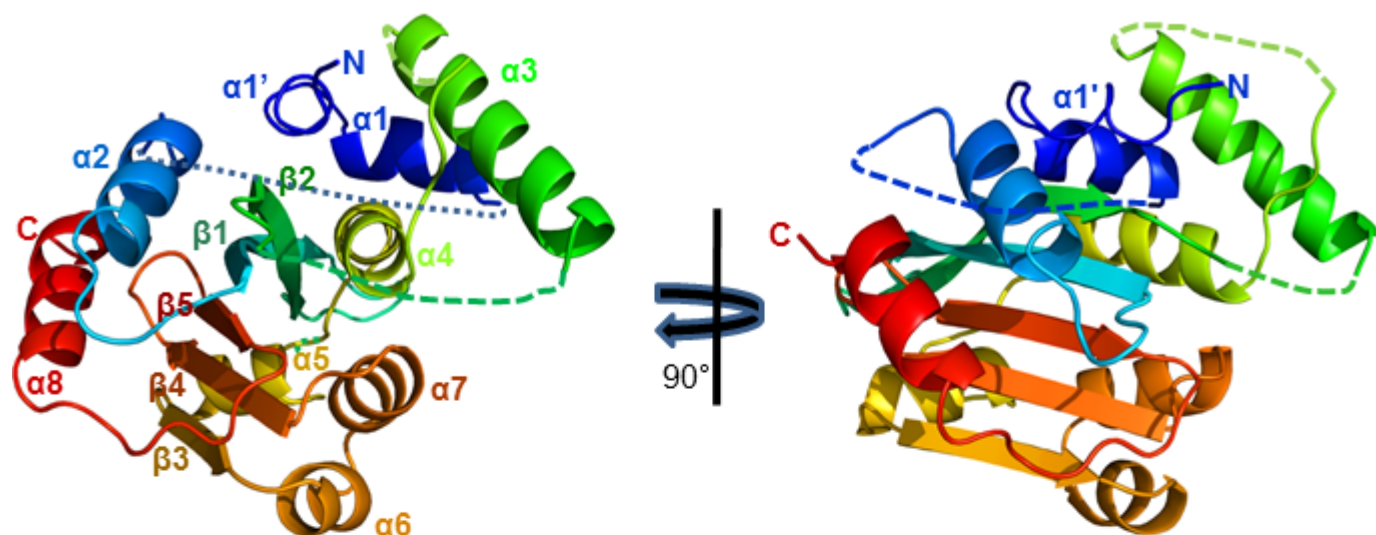
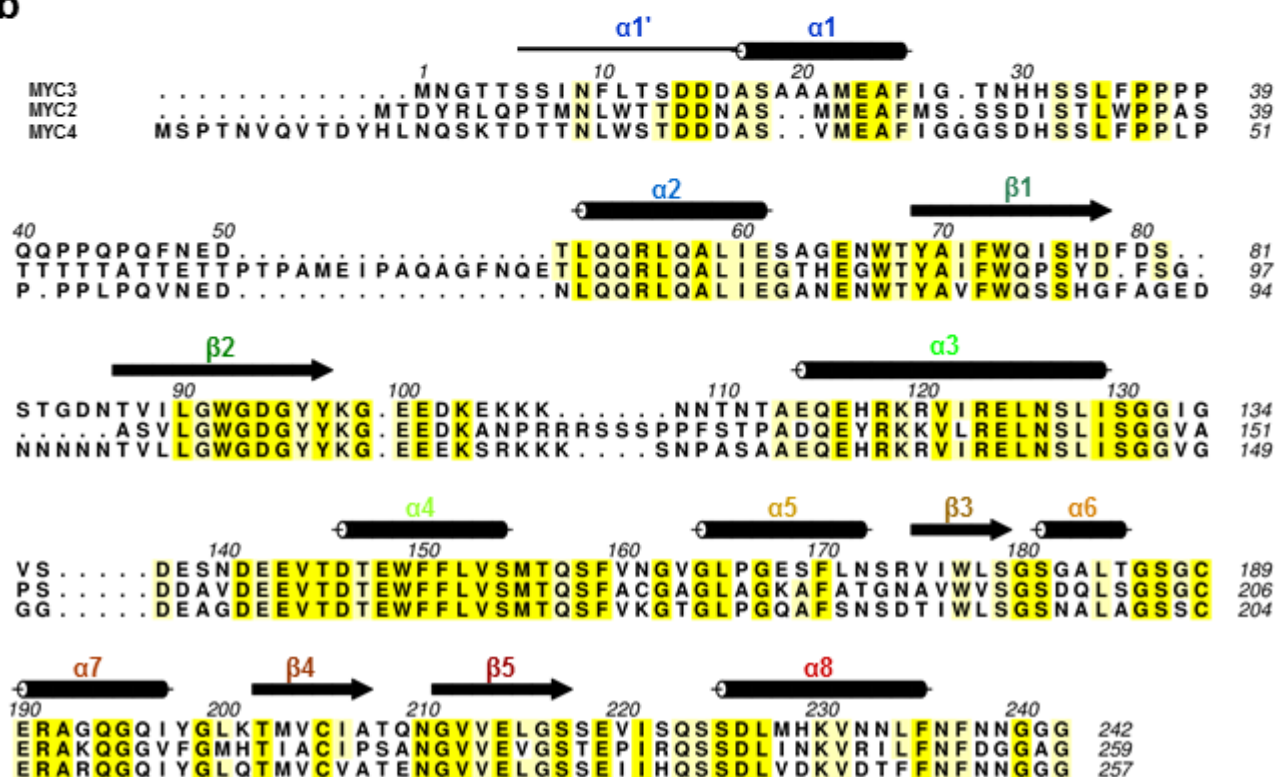
**Extended Data Figure 1 | Mapping of the JAZ9–MYC3 interface.** **a, b,** Yeast two-hybrid analysis of the interaction between LexA–JAZ9 and BD42(AD)–MYC3 constructs. Simplified diagrams of MYC3 (**a**) and JAZ9 (**b**) proteins are shown on top. Blue yeast colonies indicate a positive interaction between two proteins. The experiment was repeated three times with same results. **c,** Sequence alignment of the Jas motif of the 12 *A. thaliana* JAZ proteins. The N-terminal five amino acids that are unwound in the crystal structure of the

COI1–ligand–JAZ co-receptor complex<sup>23</sup> are indicated by a red line on top of Jas<sup>JAZ1</sup>. Asterisks denote amino acids conserved in all of the sequences, colons denote similar amino acids. **d,** Interaction between purified His6Sumo–MYC3 N-terminal proteins and biotinylated Jas motif peptides by AlphaScreen luminescence proximity assay ( $n = 3$  technical replicates; error bars, s.d.). \*\*\*Significant differences ( $P < 0.001$ ) compared with the no-Jas control by Student's *t*-test. The experiment was repeated three times with similar results.



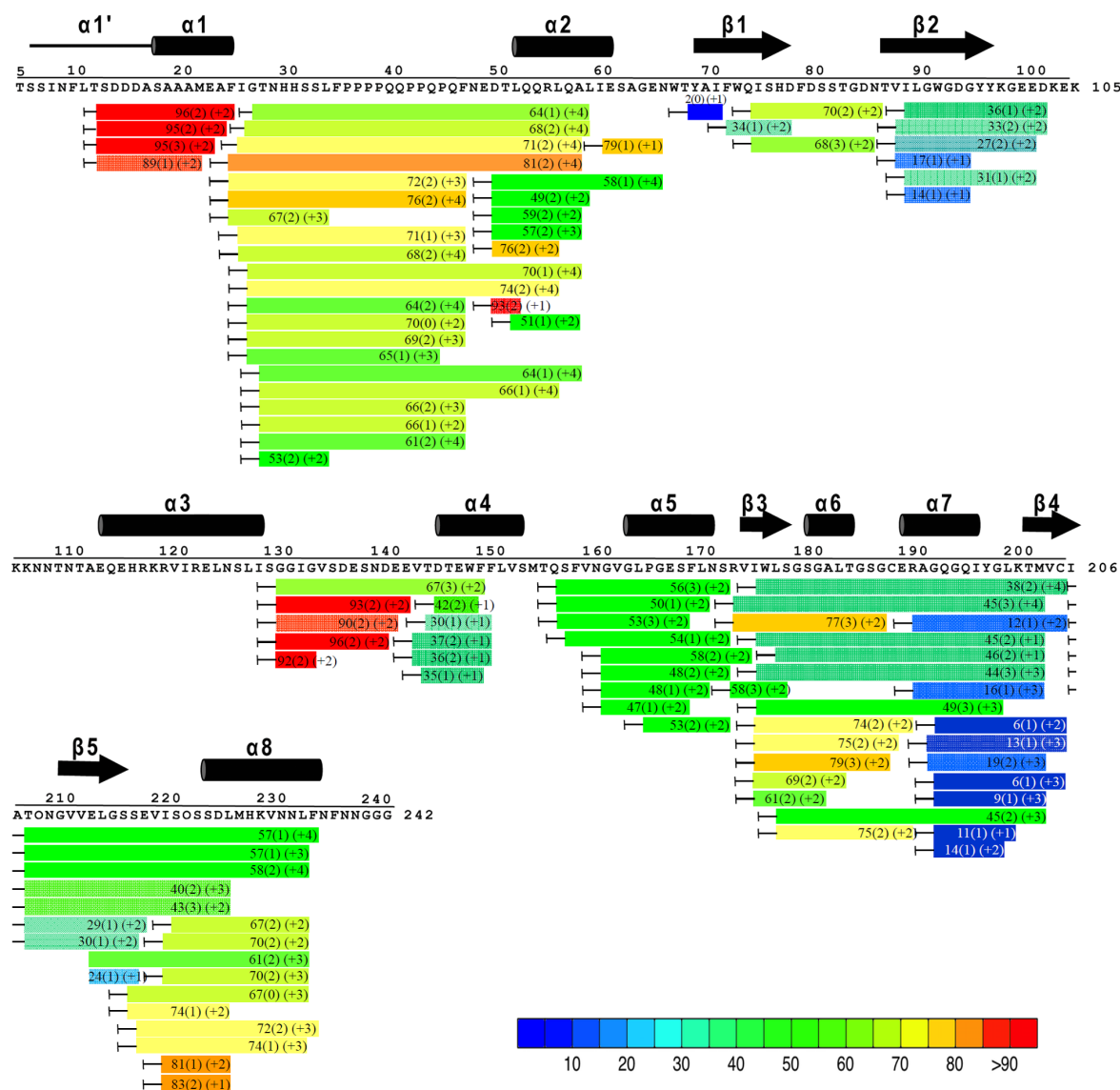
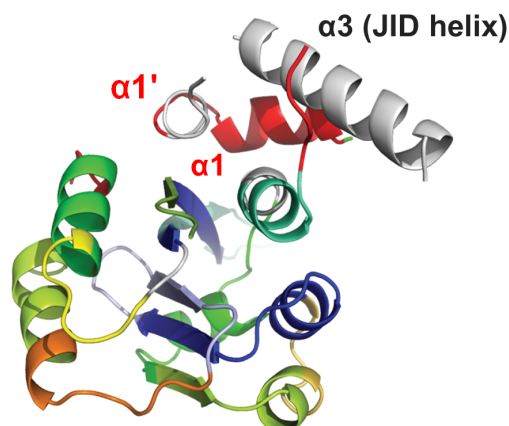
**Extended Data Figure 2 | Interactions of MYC2 and MYC4 with Jas motifs of JAZ8, JAZ9 and JAZ12.** **a**, Yeast two-hybrid assays between MYC N-terminal proteins and full-length JAZ9. The experiment was repeated three times with same results. **b**, AlphaScreen assay between His6Sumo-tagged MYC

N-terminal proteins and biotinylated JAZ peptides ( $n = 3$  technical replicates; error bars, s.d.). \*\*\*Significant differences ( $P < 0.001$ ) compared with no-Jas control by Student's  $t$ -test. The experiment was repeated three times with similar results.

**a****b**

**Extended Data Figure 3 | Arrangement of secondary structure elements in MYC3(5–242).** **a**, Rainbow colour scheme of MYC3(5–242) in two orientations, from blue (N terminus) to red (C terminus). **b**, Secondary

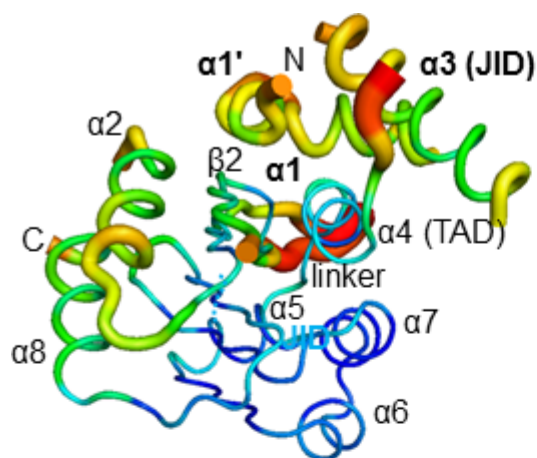
structure elements overlaid on the sequence alignment of MYC2, MYC3 and MYC4 N-terminal proteins. Note that  $\alpha 1'$  (solid line) is a loop with partial helix character and is connected to  $\alpha 1$  by a  $90^\circ$  kink.

**a****b**

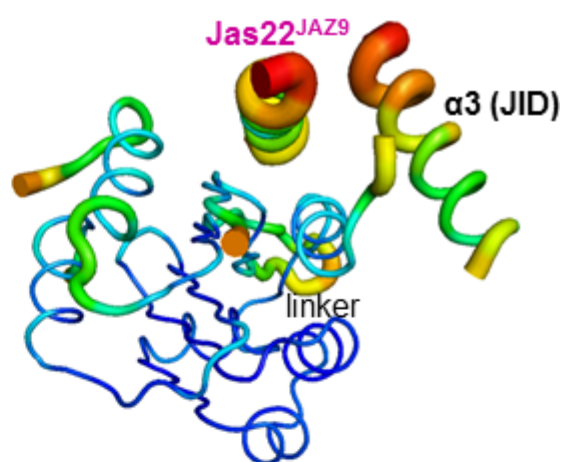
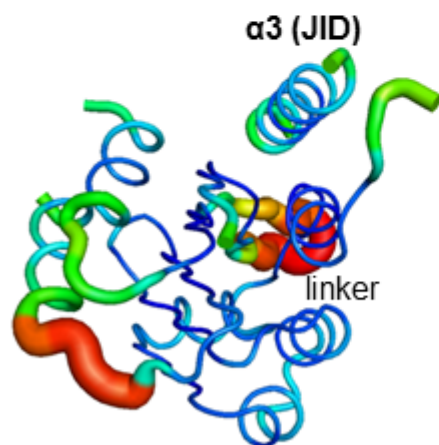
**Extended Data Figure 4 | Surface accessibility and structural dynamics of MYC3(5–242) revealed by HDX.** **a**, HDX heat map of MYC3(5–242). The colour bar indicates the percentage deuterium exchange. Three experimental repeats were performed for each HDX time point. **b**, HDX heat map

overlaid onto the MYC3(5–242) apo structure. Peptides corresponding to the JID helix were not resolved (no HDX information, grey colour), preventing a definitive assessment of the dynamics of the JID helix in solution.

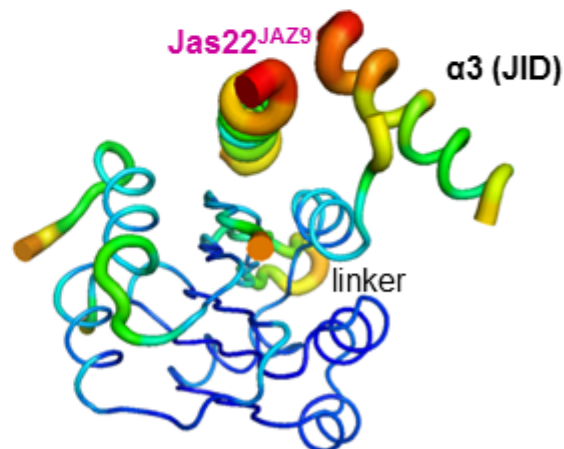




Apo MYC3(5-242)

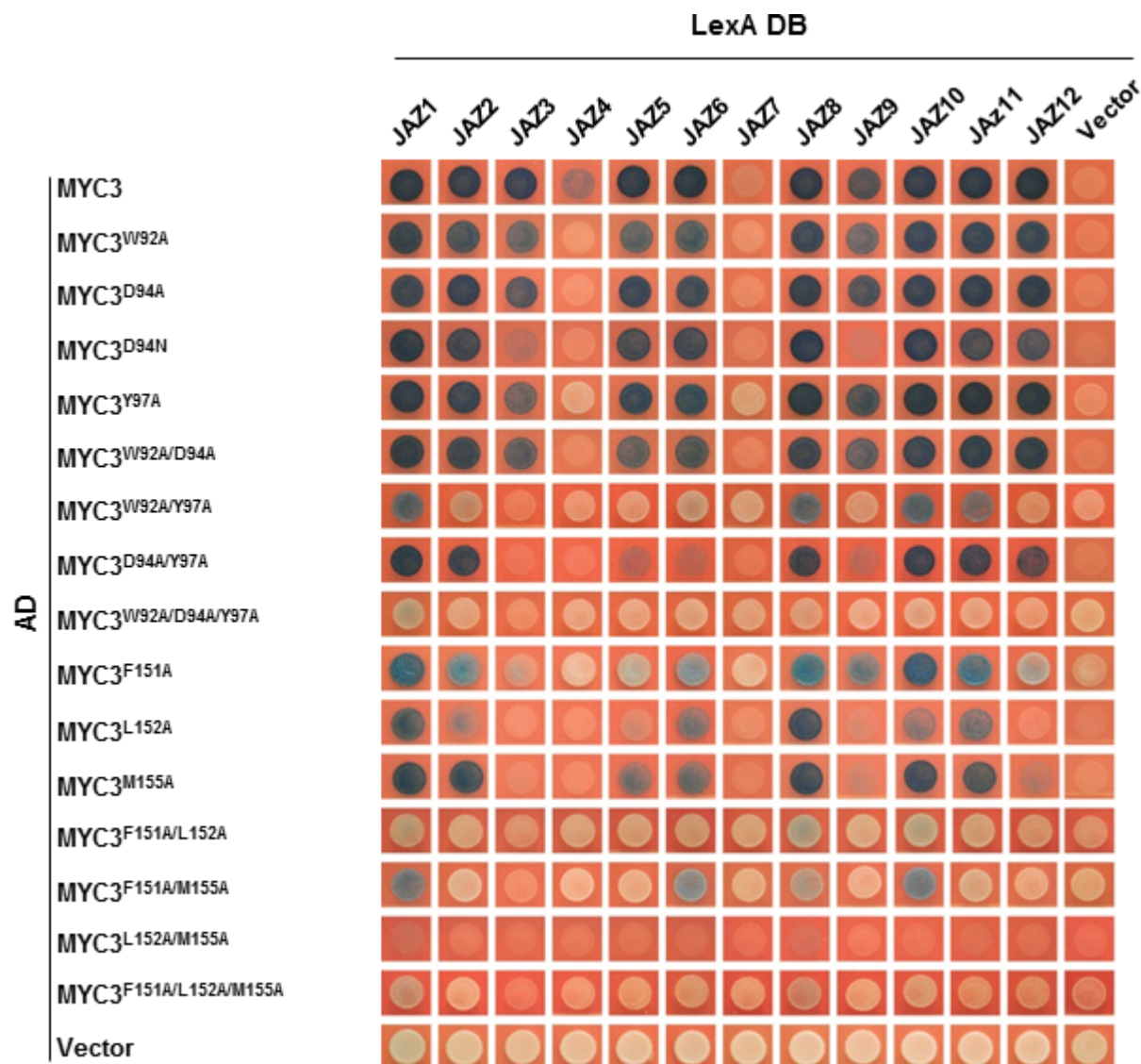
MYC3(5-242) + Jas22<sup>JAZ9</sup>

Apo MYC3(44-238)

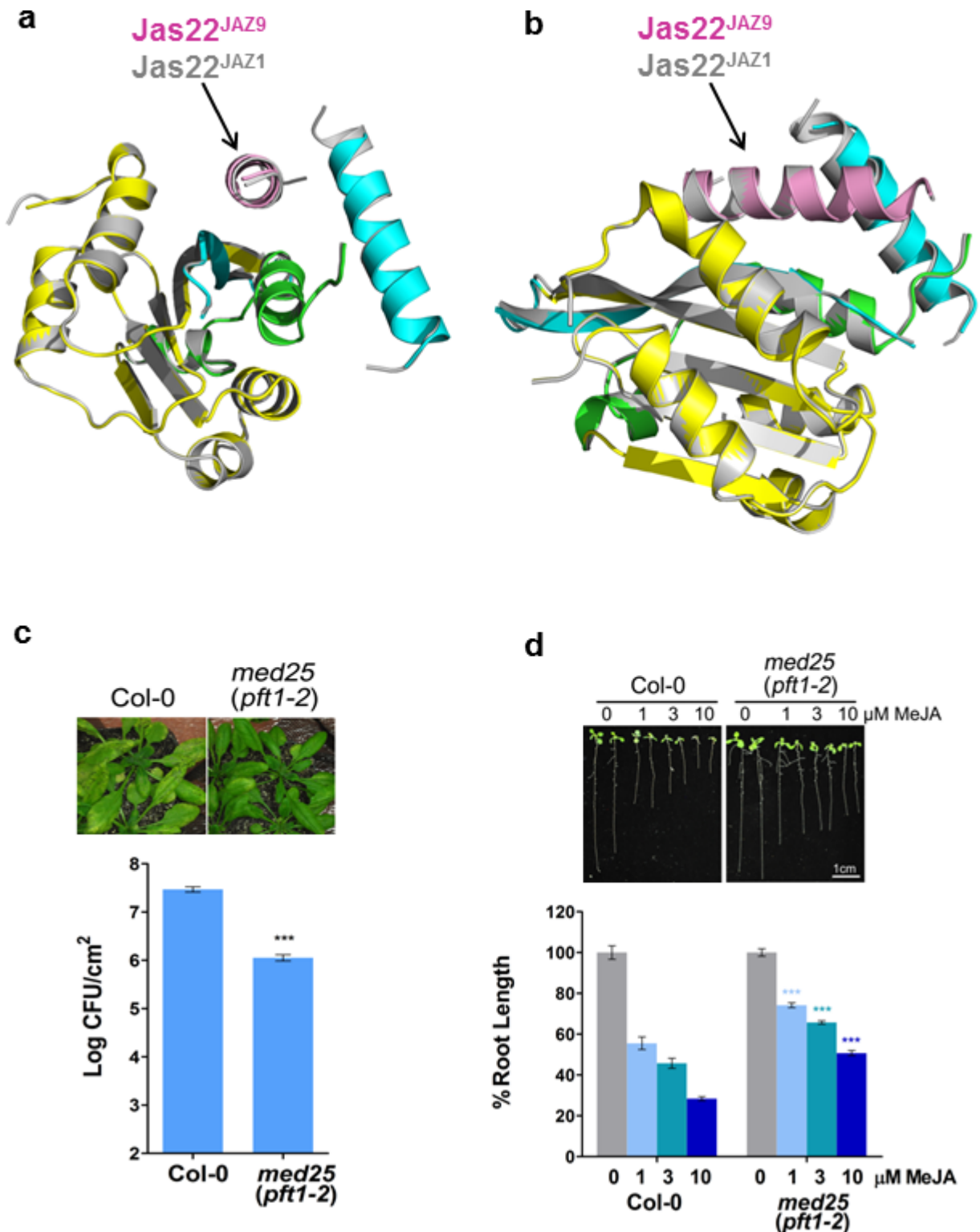
MYC3(44-238) + Jas22<sup>JAZ9</sup>

**Extended Data Figure 5 | B-factor presentations of the four crystal structures.** The B-factor indicates the dynamic mobilities of different resolved parts within the structure. The thicker the lines and the warmer the colour, the higher is the mobility. Other than two linker regions (linker), the three helices that can occupy the JID helix have the highest B-factors in all four structures. The difficulties in crystallizing the MYC3(5-242)–Jas22<sup>JAZ9</sup>

complex are therefore probably caused by its high conformational flexibility due to the presence of all three dynamic helices as well as the unfolding of the  $\alpha 1'/\alpha 1$  helix. Covalent fusion to MYC3(5-242) probably stabilizes the conformational flexibility of Jas22<sup>JAZ9</sup> and the complex. Note that the presence of the  $\alpha 1'/\alpha 1$  helix does not interfere with the ability of MYC3 to bind the JAZ peptide (compare MYC3(5-242) and MYC3(44-238) in Fig. 5b).

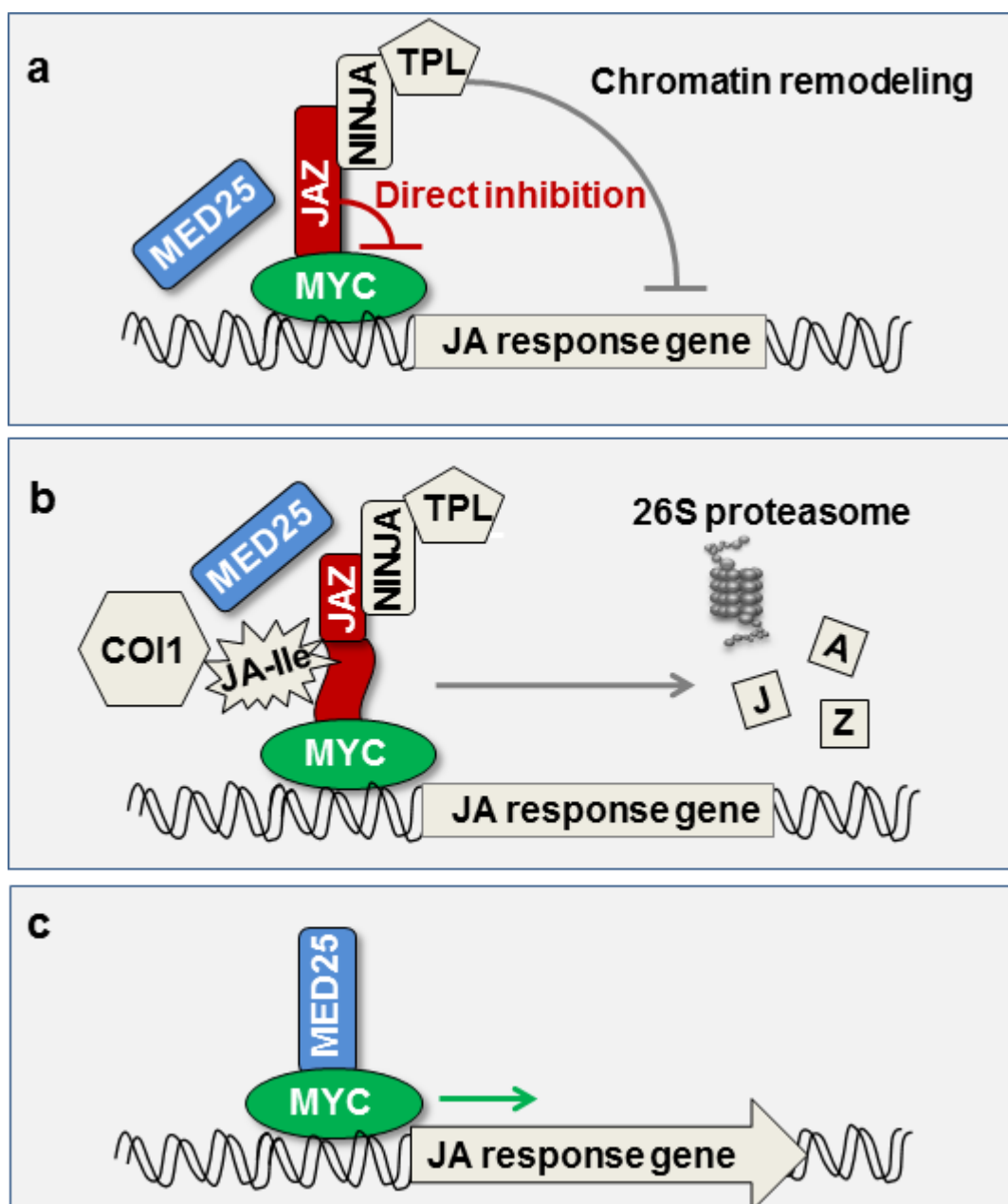


**Extended Data Figure 6 |** Effects of mutations in MYC3 on the interactions between JAZ and MYC3 proteins in yeast two-hybrid assays. The development of blue yeast colonies indicates the positive interaction between two proteins. The experiment was repeated three times with same results.



**Extended Data Figure 7 | MYC3(44–238) in complex with Jas22<sup>JAZ1</sup> and Jas22<sup>JAZ9</sup> and phenotypes of the *Arabidopsis med25* mutant.** **a, b**, MYC3(44–238) in complex with Jas22<sup>JAZ1</sup> (amino acids 200–221; grey) overlaid with MYC3(44–238) in complex with Jas22<sup>JAZ9</sup> (amino acids 218–239; pink). **c**, *Arabidopsis med25* mutant (*pft1-2*) plants are less susceptible to *P. syringae* pathovar *tomato* (*Pst*) DC3000 than *Arabidopsis* wild-type (Col-0) plants. Disease symptoms (chlorotic lesions; upper panel) and bacterial population (lower panel) of *Arabidopsis* wild-type and *med25* mutant (*pft1-2*) plants 3 days after dip-inoculation with *Pst* DC3000 at  $1 \times 10^8$  colony-forming units per millilitre ( $n = 4$  biological replicates; error bars, s.e.m.). \*\*\*Significant difference ( $P < 0.001$ ) in bacteria population between Col-0 and *med25* mutant plants, as determined by two-tailed *t*-test. The experiment was repeated five

times with similar results and the images presented are representative of five repeats. **d**, The *med25* mutant plants are less sensitive to jasmonate-induced root growth inhibition than wild-type plants. A representative picture (upper panel) and percentages of root growth inhibition (lower panel) of 10-day-old wild-type and *med25* mutant (*pft1-2*) *Arabidopsis* seedlings after treatment with 0.1% DMSO (control), 1 μM, 3 μM or 10 μM MeJA ( $n = 15$  biological replicates; error bars, s.e.m.). Triple asterisks (\*\*\*) with different colours indicate the significant differences ( $P < 0.001$ ) between Col-0 and the *med25* (*pft1-2*) mutant with the same concentration of MeJA treatment, as determined by two-way ANOVA with Bonferroni post-test. The experiment was repeated four times with similar results and the images presented are representative of four repeats.



**Extended Data Figure 8 | A simplified diagram of the core components of the jasmonate signalling cascade.** **a**, In the resting stage, jasmonate response gene expression is restrained by a family of JAZ transcriptional repressors. JAZ repressors bind and inhibit the MYC family of transcription factors either directly or through the NINJA adaptor. TPL in turn recruits histone deacetylases/methyltransferases (not shown) to repress gene expression through chromatin remodelling. **b**, In response to stress or developmental cues,

plants synthesize jasmonate-Ile, which serves as molecular glue to facilitate the formation of a co-receptor complex between JAZ and COI1. The formation of the COI1-JAZ co-receptor complex leads to ubiquitination and proteasome-dependent degradation of JAZ repressors. **c**, JAZ-free MYCs interact with the MED25 subunit of the Mediator complex and recruit RNA polymerase II (not shown) to the promoters of jasmonate-responsive genes. Components examined in this study are coloured.



Extended Data Table 1 | X-ray data collection and refinement statistics for MYC3 structures

	Native MYC3(5-242)*	SeMet MYC3(44-238)*	Native MYC3(44-238)+JAZ9 complex <sup>#</sup>	Native MYC3(44-238)+JAZ1 complex <sup>#</sup>	Native MYC3(5-242)+JAZ9 complex <sup>#</sup>
<b>PDB code</b>	4RRU	4RQW	4RS9	4YZ6	4YWC
<b>Data collection</b>					
Space group	P3 <sub>2</sub> 21	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P3 <sub>2</sub> 21	P3 <sub>2</sub> 21	C222 <sub>1</sub>
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	85.4, 85.4, 53.7	57.3, 76.6, 85.9	85.8, 85.8, 60.0	86.2, 86.2, 59.4	59.1, 110.6, 161.8
<i>a</i> , <i>b</i> , <i>g</i> (°)	90, 90, 120	90, 90, 90	90, 90, 120	90, 90, 120	90, 90, 90
Wavelength	1.0782	0.9787	0.9786	1.078	0.9787
Resolution (Å)	50-2.1	50-2.2	50-1.95	50-1.95	50-2.4
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.054 (0.97) <sup>*</sup>	0.050 (1.22) <sup>*</sup>	0.057 (1.31) <sup>*</sup>	0.049 (0.95) <sup>*</sup>	0.139 (1.08) <sup>*</sup>
<i>I</i> / <i>sI</i>	18.2 (2.2) <sup>*</sup>	20.6 (1.8) <sup>*</sup>	23.7 (2.0) <sup>*</sup>	29.7 (2.6) <sup>*</sup>	10.3 (2.0) <sup>*</sup>
Completeness (%)	100 (100) <sup>*</sup>	99.9 (99.9) <sup>*</sup>	100.0 (100.0) <sup>*</sup>	100.0 (100.0) <sup>*</sup>	99.9 (100.0) <sup>*</sup>
Redundancy	7.3 (7.4) <sup>*</sup>	8.2 (8.4) <sup>*</sup>	12.0 (10.8) <sup>*</sup>	12.2 (12.4) <sup>*</sup>	8.2 (8.3) <sup>*</sup>
<b>Refinement</b>					
Resolution (Å)	50-2.1	50-2.2	50-1.95	50-1.95	50-2.4
No. reflections	12780	18734	17922	17940	21103
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.234/0.277	0.214/0.263	0.200/0.234	0.184/0.217	0.239/0.295
No. molecules per asymmetric unit	1	2	1	1	2
No. atoms					
Protein	1447	2550	1338	1367	2729
Ligand/peptide	2	2	155	155	357
Water	71	57	137	160	100
<i>B</i> -factors					
Protein	60.0	62.5	46.9	40.3	58.8
Ligand/peptide	106.3	94.2	70.5	70.9	56.6
Water	57.4	65.4	58.3	55.9	56.8
R.m.s deviations					
Bond lengths (Å)	0.009	0.008	0.006	0.008	0.008
Bond angles (°)	1.34	1.27	1.01	1.09	1.24

<sup>#</sup>The X-ray diffraction data were obtained from a single crystal. \*Values in parentheses are for the highest-resolution shell.

Extended Data Table 2 | Main interacting residues between JAZ9 and MYC3

JAZ9 aa	Distance	Interaction	MYC3* aa
<b>R223</b>	3.1 Å 4.0 Å 4.0 Å	H-bond VdW VdW	<b>M155</b> <b>M155</b> <b>W92</b>
<b>S226</b>	2.6 Å 2.8 Å	H-bond H-bond	<b>D94</b> <b>W92</b>
<b>L227</b>	3.8 Å 4.1 Å 3.9 Å 4.2 Å	VdW VdW VdW VdW	<b>L125</b> <b>L125</b> <b>L152</b> <b>M155</b>
<b>R229</b>	2.9 Å 3.4 Å	Ionic ionic	<b>D94</b> <b>D94</b>
<b>F230</b>	3.9 Å 4.5 Å 4.1 Å 3.8 Å 3.9 Å	VdW VdW VdW VdW VdW	<b>Y97</b> <b>Y97</b> <b>E148</b> <b>F151</b> <b>F151</b>
<b>L231</b>	4.1 Å 4.0 Å 4.5 Å 4.4 Å	VdW VdW VdW VdW	<b>L125</b> <b>N126</b> <b>I122</b> <b>I122</b>
<b>K233</b>	3.2 Å 3.4 Å	H-bond H-bond	<b>Y97</b> <b>Y96</b>
<b>R234</b>	2.9 Å 3.3 Å 2.9 Å 3.0 Å 3.0 Å	H-bond H-bond Ionic Ionic H-bond	<b>E143</b> <b>E143</b> <b>E148</b> <b>E148</b> <b>E142</b>
<b>K235</b>	3.0 Å	H-bond	<b>N126</b>

\*Green, TAD amino acid; cyan, JID amino acid; bold, residues whose mutation to alanine compromised the JAZ-MYC interaction in yeast two-hybrid and AlphaScreen assays. VdW, Van der Waals bond.

# Real-time observation of the initiation of RNA polymerase II transcription

Furqan M. Fazal<sup>1\*</sup>, Cong A. Meng<sup>2\*</sup>, Kenji Murakami<sup>3†\*</sup>, Roger D. Kornberg<sup>3</sup> & Steven M. Block<sup>1,4</sup>

Biochemical and structural studies have shown that the initiation of RNA polymerase II transcription proceeds in the following stages: assembly of the polymerase with general transcription factors and promoter DNA in a 'closed' preinitiation complex (PIC)<sup>1,2</sup>; unwinding of about 15 base pairs of the promoter DNA to form an 'open' complex<sup>3,4</sup>; scanning downstream to a transcription start site; synthesis of a short transcript, thought to be about 10 nucleotides long; and promoter escape. Here we have assembled a 32-protein, 1.5-megadalton PIC<sup>5</sup> derived from *Saccharomyces cerevisiae*, and observe subsequent initiation processes in real time with optical tweezers<sup>6</sup>. Contrary to expectation, scanning driven by the transcription factor IIH<sup>7–12</sup> involved the rapid opening of an extended transcription bubble, averaging 85 base pairs, accompanied by the synthesis of a transcript up to the entire length of the extended bubble, followed by promoter escape. PICs that failed to achieve promoter escape nevertheless formed open complexes and extended bubbles, which collapsed back to closed or open complexes, resulting in repeated futile scanning.

Optical tweezers have been used in studies of transcript elongation by RNA polymerase II (Pol II)<sup>13–15</sup> with the use of a 'dumbbell' configuration, consisting of two beads held in separate optical traps, connected by a segment of DNA. One bead was directly attached to Pol II, and the other was attached to the opposite end of the template DNA, minimally around 3 kilobases (kb), for traversing the distance between the traps. For the study of transcription initiation, we adapted a similar approach to the PIC. Pol II, biotinylated for attachment to one bead, was assembled together with transcription factors on DNA, and end-labelled with digoxigenin for attachment to the other bead. The transcription factors consisted of six general transcription factors (GTFs) including TATA-binding protein (TBP), transcription factor IIB (TFIIB), TFIIE, TFIIH and TFIIA and Sub1 (yeast homologue of human PC4), which is thought to stabilize the PIC<sup>16</sup> (Extended Data Fig. 1). The DNA contained the *SNR20* (also known as *LSR1*) promoter fused to an additional 2.7 kb length of DNA, sufficient to separate the beads by roughly the wavelength of light. The *SNR20* promoter bore a mutation resulting in one, rather than many, transcription start sites (TSSs). Two versions of the promoter were used: the otherwise wild-type promoter with the single TSS located 91 base pairs (bp) downstream of the TATA box (referred to as *SNR20\** long), and a deleted version, in which the TSS was situated 31 bp downstream (*SNR20\** short), a distance characteristic of metazoan transcription. Both versions of the promoter have been characterized in bulk transcription experiments<sup>17</sup>. The PIC, assembled without the peripheral component TFIK<sup>18,19</sup>, was mixed with a 25-fold molar excess of PIC without the additional 2.7 kb DNA, to achieve an overall PIC concentration sufficient to avoid dissociation. A twofold excess of TFIK was added, and dumbbells were formed by reaction of the PIC mixture with anti-digoxigenin-coated and avidin-coated beads (Extended Data Fig. 2).

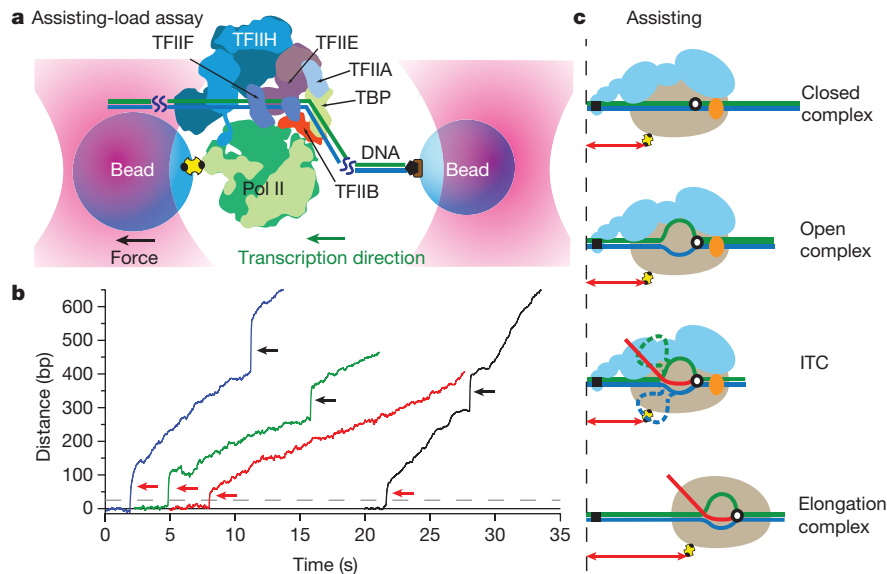
In a dumbbell carrying digoxigenin on the upstream end of the DNA, the tension exerted by the optical trap tends to pull the polymerase downstream, in the same direction as transcription, resulting in an 'assisting-load' assay (Fig. 1a). Transcription was initiated by the addition of saturating concentrations of all four ribonucleoside triphosphates (rNTPs). Force was maintained during measurements by the use of an optical force clamp, as the location of the polymerase on DNA was tracked with sub-nanometre-level precision. Transcription was signalled by movement of the polymerase (Fig. 1b, *SNR20\** short promoter) at  $29 \pm 3 \text{ bp s}^{-1}$  ( $n = 10$ , mean  $\pm$  s.e.m.), consistent with elongation rates observed in previous assays of transcription under similar assisting loads<sup>14,15</sup>. To confirm the identification of the moving molecules as transcription elongation complexes, we raised the force instantaneously to a value (10–15 pN) that, in our experience, can only be sustained by a stable elongation complex (Fig. 1b, black arrows). Only 2–3% of dumbbells gave rise to transcription elongation complexes, whereas in biochemical assays, about 18% of PICs gave rise to runoff transcripts (Extended Data Fig. 3a). The lower efficiency of initiation in the single-molecule system was attributable to the much lower protein concentrations used ( $<1 \text{ nM}$ , at least tenfold lower than biochemical assays; Extended Data Fig. 3b).

The onset of polymerase movement at a rate characteristic of transcript elongation was preceded by an almost instantaneous jump (Fig. 1b, red arrows), occurring around  $15 \pm 2 \text{ s}$  ( $n = 10$ ) after the addition of rNTPs. No such movement was observed in the absence of rNTPs. An interpretation consistent with all other available information is that the polymerase draws downstream DNA into the active centre region to form an extended unwound region, or transcription bubble, and then lurches forward after DNA rewinding and bubble collapse (Fig. 1c). Because one bead is attached to the upstream end of the DNA and the other bead to the polymerase, there is no change in the distance between them when DNA is drawn in from the downstream side. Only once the polymerase is released from its point of attachment at the upstream edge of the bubble (promoter escape), and DNA rewinds to collapse the bubble, does the distance between the beads change and lengthen (Fig. 1c). The size of the jump at the transition to a transcription elongation complex was  $70 \pm 13 \text{ bp}$  ( $n = 9$ , mean  $\pm$  s.e.m.), with a minimum of 32–34 bp and a maximum of about 140 bp.

The jump after promoter release and the corresponding transition to a stable elongation complex are notable in two further respects. First, the bubble does not collapse completely at the jump, because about 15 bp remain unwound in the Pol II active centre as a transcription bubble from the time of open complex formation until the end of transcript elongation<sup>20</sup>. Therefore, the entire length of the unwound region in this initial transcribing complex (ITC) is, on average, approximately 85 bp ( $70 + 15 \text{ bp}$ ). Second, because this experiment was performed with the *SNR20\** short promoter (in which the TSS is located 31 bp downstream of the TATA box), transcription was initiated within the open complex, and the nascent transcript extended

<sup>1</sup>Department of Applied Physics, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Department of Chemistry, Stanford University, Stanford, California 94305, USA. <sup>3</sup>Department of Structural Biology, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Department of Biology, Stanford University, Stanford, California 94305, USA. <sup>†</sup>Present address: Department of Biochemistry and Molecular Biophysics at the Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

\*These authors contributed equally to this work.



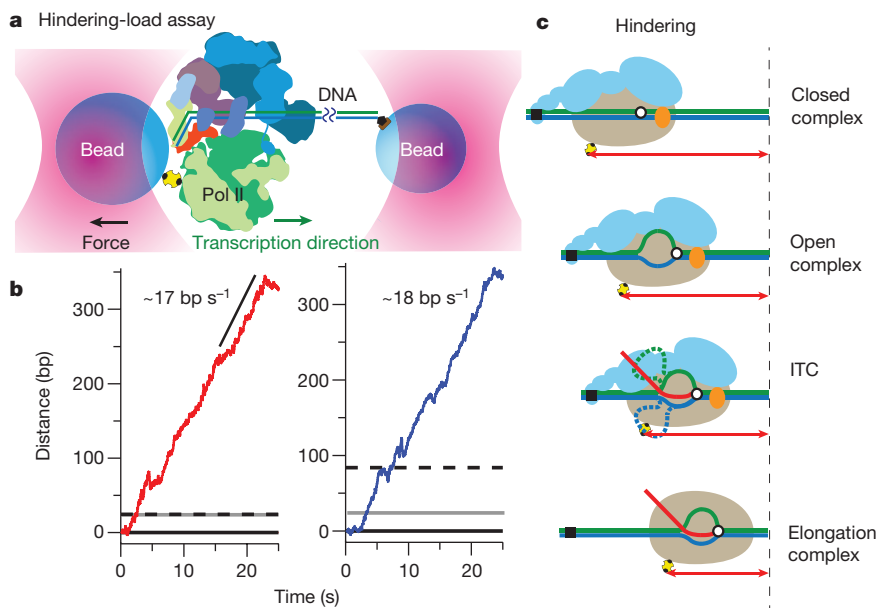
**Figure 1 | Transcription initiation in assisting-load assay.** **a**, A dumbbell tether (not to scale; subunit Sub1 not shown) is formed between beads (blue) held in separate optical tweezers (pink), with one attached to Pol II (green) in the PIC via an avidin–biotin linkage (yellow, black), and the other to upstream DNA via a digoxigenin linkage (brown, black). As transcription proceeds (green arrow indicates direction), the tether extension increases. **b**, Representative records of Pol II elongation (dashed grey line denotes the TSS location) after promoter escape (red arrows), with the applied force often stepped up after ~10 s to confirm elongation (black arrows; the associated

vertical discontinuity is due to tether stretch). **c**, Ssl2 subunit in TFIIH (orange) unwinds the template (blue) and non-template (green) strands of DNA around the active site (white circle, black outline) of Pol II (beige), and creates a transcription bubble (open complex formation). RNA synthesis while still bound to the promoter results in DNA scrunching at the upstream edge of transcription bubble (ITC), which re-anneals after Pol II enters productive elongation (elongation complex). Distances measured by the assays are indicated (double-headed red arrows). Not to scale.

to the downstream end of the unwound region before the jump (Fig. 1c). Evidently, a transcript averaging 70 nucleotides, and as long as around 140 nucleotides, is synthesized before promoter clearance and the transition to a stable elongation complex.

To observe the process occurring before the jump, presumed to involve the drawing of downstream DNA into the Pol II active centre, we moved the point of attachment of the bead to DNA from the upstream to the downstream end of the template. In this configuration, external forces applied to polymerase tend to pull it upstream, opposite to the direction of transcription, resulting in a ‘hindering-load’ assay (Fig. 2a). Fewer dumbbells (<2%) yielded transcription elongation complexes in this assay, consistent with previous studies showing that

hindering loads reduce polymerase processivity. Dumbbells that elongated did so at  $17\text{--}18\text{ bp s}^{-1}$  (Fig. 2b), consistent with previous measurements of Pol II transcription under hindering loads<sup>14,15</sup>. In contrast to the assisting-load assay, there was no jump at the transition to a stable elongation complex, but rather a gradual distance change (Fig. 2b). The distance change was the same size as the jump in the assisting-load assay, and was observed for both forms of the *SNR20\** promoter. In the case of the *SNR20\** long promoter, which initiates transcription downstream (Fig. 2, dashed line), the distance change reflects open complex formation and scanning to the TSS; in the case of the *SNR20\** short promoter, which initiates transcription in the open complex (dashed line, Fig. 2b), the distance change reflects open



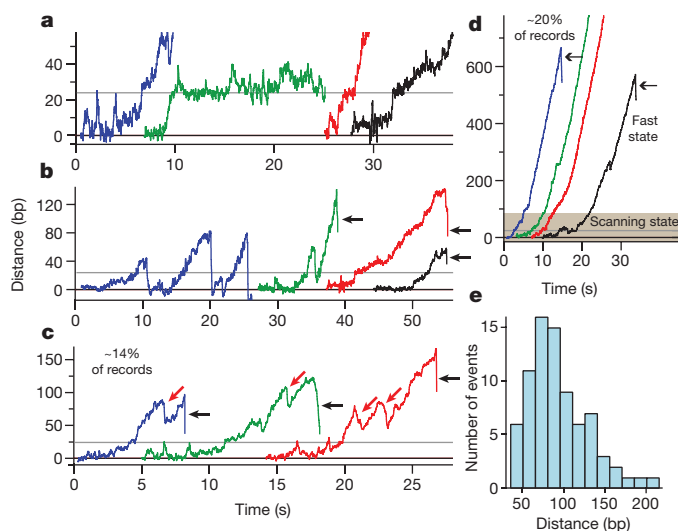
**Figure 2 | Transcription initiation in hindering-load assay.** **a**, By attaching one bead to downstream DNA, a hindering-load assay was developed (not to scale, Sub1 not shown). Scanning and subsequent transcription events (green arrow indicates transcription direction) resulted in tether-extension decrease. **b**, Records illustrating Pol II escape and elongation, with a velocity of  $\sim 17\text{--}18\text{ bp s}^{-1}$ , collected on *SNR20\** short (red, left panel) and *SNR20\** long (blue, right panel) in the presence of rNTPs. The dashed black line denotes the TSS at +1; the solid grey line marks position of the predicted  $\sim 24\text{ bp}$  distance change after open complex formation. **c**, Distances measured by the assays are indicated (colour scheme same as Fig. 1c).



complex formation, extension of the bubble, and transcription in the ITC (Fig. 2c).

In the single-molecule system, it was possible to investigate not only PICs that initiated transcription, but also those that failed to do so. Approximately 20% of dumbbells showed movement downstream without the initiation of transcription (Fig. 3 and Extended Data Fig. 4). The downstream movement began with an initial distance change of about 24 bp, which was often punctuated, at the temporal resolution of our assay ( $\sim 0.1$  s), by brief ( $< 2$  s) pauses (Fig. 3a, b). After the initial 24 bp, movement continued to a maximum of about 150–200 bp downstream (Fig. 3e), until either the bubble collapsed back to a distance of 24 bp or 0 bp, or the PIC dissociated, as evidenced by rupture of the dumbbell (Fig. 3b, black arrows). Bubble collapse was often followed by a repetition of the downstream movement.

The downstream movement was processive, and was observed under all three conditions examined: SNR20\* short with rNTPs ( $n = 40$ ) (Fig. 3a–d); SNR20\* long with rNTPs ( $n = 19$ ) (Extended Data Fig. 4a); and SNR20\* long with dATP ( $n = 15$ ) (Extended Data Fig. 4b). No movement was observed in the absence of rNTPs or dATP. There were no significant differences in either processivity or velocity in the three conditions. Combining these data yielded a pause-free velocity for downstream movement of  $36 \pm 1$  bp s $^{-1}$  ( $n = 24$ ). Because the velocity was unchanged when only dATP was present (and no rNTPs), it must have been produced by TFIIF activity, and not by polymerase. The extent of the downstream movement was  $94 \pm 36$  bp (mean  $\pm$  s.d.) (Fig. 3e). TFIIF, which contains a kinase responsible for phosphorylation of the carboxy-terminal domain of Pol II (refs 18, 19), could be omitted without effect. There was no change in the absence of TFIIF in either the distance ( $92 \pm 33$  bp,  $n = 34$ ) or the velocity ( $36 \pm 2$  bp s $^{-1}$ ,  $n = 4$ ) of downstream movement. In about 20% ( $n = 15$  out of 74) of dumbbells that displayed TFIIF activity but failed to initiate transcription, there was a transition to a ‘fast state’, characterized by a velocity of  $61 \pm 2$  bp s $^{-1}$  and downstream movement through hundreds of base pairs (Fig. 3d). The transition to the fast state was irreversible, and must reflect action of the TFIIF helicase subject to little or no restraint by other GTFs or Pol II.



**Figure 3 | Records of TFIIF motion for the SNR20\* short construct with rNTPs present in hindering-load assay.** **a**, Initial transition from the closed (0 bp) to open ( $\sim 24$  bp predicted distance change, grey line) complex. **b**, Scanning behaviour, with occasional bubble collapse to the closed complex (blue record) or open complex (green record). **c**, Infrequent slips in the records ( $n = 10$  of 74) were observed (red arrows). **d**, Occasional irreversible transition from scanning (shaded region) to a highly processive fast state, occurring at a distance of  $130 \pm 21$  bp ( $n = 9$ , mean  $\pm$  s.e.m.). In all records, black arrows mark tether breakage, probably due to PIC dissociation. **e**, Histogram of TFIIF processivity, with a peak between 40 and 140 bp ( $n = 78$ ).

Biochemical evidence for movement driven by the TFIIF helicase was obtained by exonuclease III footprinting of the PIC (Extended Data Fig. 5). Extended regions of unwound DNA were previously revealed by KMnO $_4$  reactivity of yeast and *Drosophila* promoters in studies of transcriptional activity *in vivo*<sup>21–23</sup>. The regions were similar in all cases, extending from about 20–60 bp downstream of the TATA box, with the TSSs of the *Drosophila* promoters near the upstream edge of the unwound region nearest the TATA box, and the TSSs of the yeast promoters near the downstream edge furthest from the TATA box. It had been thought that transcribing polymerases with 15-bp bubbles, at different locations on individual promoters and then revealed collectively in the KMnO $_4$  analysis, gave the appearance of an extended bubble. Our examination of single molecules suggests instead that an extensive unwound region is a characteristic of every individual promoter, rather than some collective property. We obtained similar results for a yeast promoter in which the TSS was located near the downstream edge of the unwound region, and for the same promoter in which the TSS was moved to the upstream edge, as in *Drosophila* and other metazoans.

The formation of the unwound region is not a consequence of transcription, but instead of TFIIF action, because it occurs in the presence of dATP without rNTPs, and because it is observed even when the TSS is located at the downstream end of the unwound region. TFIIF must act continuously to maintain the unwound region, consistent with previous biochemical studies showing a requirement of TFIIF to prevent premature arrest of ITCs<sup>24,25</sup>. It is not known what determines the length of the unwound region, nor where the unwound DNA resides in the complex. In the case of bacterial polymerase transcription, approximately 10 bp of DNA drawn into the active centre before the transition to elongation are thought to be accommodated by a ‘scrunching’ mechanism<sup>26,27</sup>. The possibility of scrunching in eukaryotic transcription has previously been considered<sup>28</sup>, but no evidence obtained. Because of rotation in the direction of unwinding by TFIIF, there is unlikely to be associated torsional strain, as presumed to occur in the bacterial system. The location and conformational state of the approximately 85 bp of DNA unwound in the Pol II PIC thus remain open questions.

Although most PICs ( $\sim 80\%$  of PICs in biochemical assay conditions, and 97–98% of PICs under single-molecule assay conditions) fail to yield stable elongation complexes, they are not inert. About 20% of the dumbbells showed downstream movements of polymerase along DNA in the hindering-load assay. There was often an initial movement of  $24 \pm 2$  bp (mean  $\pm$  s.d.), which we attribute to the formation of an open complex, on the basis of previous studies of the PIC<sup>29</sup> (see Methods).

After open complex formation, downstream movement continued for a total distance of 94 bp, on average, with bubble collapse back to the open or closed complexes and repetition of downstream movement, before either final dissociation of the PIC or rupture of the dumbbell (Extended Data Fig. 6). The movement of 94 bp in the hindering-load assay is noteworthy for two reasons. First, it is in excellent agreement with the results of the assisting-load assays, in which a jump of about 70 bp was observed before the onset of transcription elongation. This jump is attributed to collapse of an extended bubble, leaving the original open complex in place, and  $70 + 24$  bp (open complex) = 94 bp. Second, the distance of downstream movement in the hindering-load assay ranged from about 30 to 150 bp, that is, about 37–157 bp from the TATA box, similar to the distribution of TSSs in yeast, which are located 40–120 bp downstream from the TATA box. Therefore, downstream movements in the hindering-load assay may be attributed to TSS scanning, which precedes the onset of transcription elongation, as observed in the assisting-load assay, and the initiation of transcription in yeast *in vivo*<sup>24,25</sup>.

It is commonly noted that TSSs for yeast promoters are spread over a wide region, rather than concentrated near the TATA box, as in metazoans. Nevertheless, as discussed above, our evidence for an extended bubble in the yeast PIC corresponds well with extended bubbles mapped

by  $\text{KMnO}_4$  reactivity in *Drosophila*<sup>21</sup>. Moreover, TATA-less promoters, which predominate in metazoans as well as in yeast, have several TSSs spread over regions of 50–100 bp in human cells<sup>30</sup>. Our findings from single-molecule studies of the yeast PIC are therefore likely to hold true for other eukaryotes as well, including metazoans.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 10 February; accepted 3 July 2015.**

**Published online 2 September 2015.**

- Conaway, R. C. & Conaway, J. W. General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190 (1993).
- Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proc. Natl Acad. Sci. USA* **104**, 12955–12961 (2007).
- Holstege, F. C., Fiedler, U. & Timmers, H. T. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* **16**, 7468–7480 (1997).
- Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005).
- Murakami, K. *et al.* Formation and fate of a complete, 31-protein, RNA polymerase II transcription initiation complex. *J. Biol. Chem.* **288**, 6325–6332 (2013).
- Fazal, F. M. & Block, S. M. Optical tweezers study life under tension. *Nature Photon.* **5**, 318–321 (2011).
- Sawadogo, M. & Roeder, R. G. Energy requirement for specific transcription initiation by the human RNA polymerase II system. *J. Biol. Chem.* **259**, 5321–5326 (1984).
- Schaeffer, L. *et al.* DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260**, 58–63 (1993).
- Svejstrup, J. Q. *et al.* Different forms of TFIIF for transcription and DNA repair: holo-TFIIF and a nucleotide excision repairosome. *Cell* **80**, 21–28 (1995).
- Dvir, A. *et al.* A role for ATP and TFIIF in activation of the RNA polymerase II preinitiation complex prior to transcription initiation. *J. Biol. Chem.* **271**, 7245–7248 (1996).
- Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1421 (2000).
- Grünberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Struct. Mol. Biol.* **19**, 788–796 (2012).
- Galbur, E. A. *et al.* Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446**, 820–823 (2007).
- Larson, M. H. *et al.* Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl Acad. Sci. USA* **109**, 6555–6560 (2012).
- Schweikhard, V. *et al.* Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proc. Natl Acad. Sci. USA* **111**, 6642–6647 (2014).
- Sikorski, T. W. *et al.* Sub1 and RPA associate with RNA polymerase II at different stages of transcription. *Mol. Cell* **44**, 397–409 (2011).
- Murakami, K. *et al.* Uncoupling promoter opening from start-site scanning. *Mol. Cell* **59**, 133–138 (2015).
- Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIK. *Cell* **79**, 1103–1109 (1994).
- Roy, R. *et al.* The MO15 cell cycle kinase is associated with the TFIIF transcription-DNA repair factor. *Cell* **79**, 1093–1101 (1994).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Giardina, C., Perez-Riba, M. & Lis, J. T. Promoter melting and TFIID complexes on *Drosophila* genes *in vivo*. *Genes Dev.* **6**, 2190–2200 (1992).
- Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
- Giardina, C. & Lis, J. T. Dynamic protein–DNA architecture of a yeast heat shock promoter. *Mol. Cell Biol.* **15**, 2737–2744 (1995).
- Dvir, A., Conaway, R. C. & Conaway, J. W. A role for TFIIF in controlling the activity of early RNA polymerase II elongation complexes. *Proc. Natl Acad. Sci. USA* **94**, 9006–9010 (1997).
- Spangler, L., Wang, X., Conaway, J. W., Conaway, R. C. & Dvir, A. TFIIF action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA. *Proc. Natl Acad. Sci. USA* **98**, 5544–5549 (2001).
- Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
- Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
- Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature Struct. Mol. Biol.* **13**, 603–610 (2006).
- Murakami, K. *et al.* Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* **342**, 1238724 (2013).
- Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Rev. Genet.* **8**, 424–436 (2007).

**Acknowledgements** We thank A. Chakraborty and B. Milic for careful reading of the manuscript, P.-J. Mattei for help with protein purification, and R. Landick and J. Gelles for discussions. This research was supported by NIH grants GM36659 and AI21144 to R.D.K. and GM57035 to S.M.B., and an NSF graduate fellowship to F.M.F.

**Author Contributions** F.M.F., C.A.M. and K.M. designed the experiments with input from, and supervision by, S.M.B. and R.D.K. F.M.F. and C.A.M. collected the single-molecule data. K.M. purified and reconstituted the PIC components, and performed the bulk experiments. F.M.F. analysed the single-molecule data, and C.A.M. and K.M. carried out the modelling. F.M.F., C.A.M., K.M., S.M.B. and R.D.K. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.B. ([sblock@stanford.edu](mailto:sblock@stanford.edu)) or R.D.K. ([kornberg@stanford.edu](mailto:kornberg@stanford.edu)).

## METHODS

**Single-molecule optical-trapping assay.** The 29-subunit yeast PIC containing biotinylated Pol II was assembled on SNR20\* promoter DNA fused to a 2.7 kb DNA 'handle'. The DNA handle allowed us to form tethers in both the hindering-load and assisting-load assays by incorporating a digoxigenin tag via PCR at either the downstream or upstream end of the DNA, respectively. The constructs containing the handle were mixed in a 1:25 molar ratio with identical PICs assembled on the same promoter DNA, but without the handle, such that the overall concentration of the PIC was 100 nM. PIC complexes assembled on DNA lacking a handle sequence are unable to form tethers, and instead serve to increase the overall concentration of PIC by mass action. The resulting mixture was incubated with twofold excess of TFIIF<sup>18,19,31,32</sup> at room temperature for 20 min to form the complete 32-subunit PIC. This complete PIC was incubated with both anti-digoxigenin-coated 0.9-µm diameter beads and avidin-coated 0.6-µm diameter beads, resulting in tethers being formed with the digoxigenin-containing handle at one end and the biotinylated Pol II in the PIC at the other end. On completion of this step, the concentration of PIC was ~25 nM. In this latter step (and all subsequent steps), the buffer used (50 mM HEPES, pH 7.5, 80 mM potassium acetate, 5 mM MgSO<sub>4</sub>, 10 mM dithiothreitol (DTT), 10% glycerol) was always supplemented with 250 nM TFIIB and 250 nM TFIIE to stabilize and maintain the PIC. The assembled dumbbell tethers were flowed into a ~5 µl flow chamber, and rinsed with ~10 µl of additional buffer to remove excess beads. Each bead that formed the tether was held in a separate optical trap, allowing controlled load to be applied on the dumbbell by using an active force clamp as previously described<sup>14</sup>. Force uncertainties were estimated at ~15% owing to variations in bead size and systematic calibration errors. The temperature on the trap was estimated to be 26 ± 1 °C (mean ± s.d.)<sup>33</sup>. Single tethers were identified as described<sup>33</sup>, and held at ~4 pN constant force for ~15–20 s, after which transcription buffer (50 mM HEPES, pH 7.5, 80 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 10% glycerol, 1 U of RNaseOUT (Life Technologies), 250 nM TFIIB and 250 nM TFIIE) containing either 1.6 mM (2×) NTPs or 1.6 mM (2×) dATP, was flowed into the flow cell while holding the tether at the same force. In the absence of nucleotides, the dumbbells could be held at about 4 pN without breakage for extended periods. An oxygen-scavenging system (8.3 mg ml<sup>-1</sup> glucose (Sigma), 46 U ml<sup>-1</sup> glucose oxidase (Calbiochem), 94 U ml<sup>-1</sup> catalase (Sigma)) was used to reduce photodamage. Data were collected at 2 kHz sampling frequency, filtered at 1 kHz with an 8-pole Bessel filter (Krohn-Hite) and boxcar averaged over a 20-point window to provide positional feedback to an active force clamp at a rate of 100 Hz. The resulting data was analysed using Igor Pro (Wavemetrics).

**Protein purification.** TFIIA, TFIIB and TBP were expressed in bacteria, and TFIIE, TFIIF and TFIH were isolated from yeast<sup>5,17,29,34</sup>. Biotinylated Pol II was isolated as previously published<sup>15</sup>. For the expression of recombinant Sub1 (refs 16, 35–37), the *Escherichia coli* Rosetta2 (DE3) strain (Stratagene) was transformed with pCold II vector (Clontech) containing the *SUB1* (also known as *TSP1*) gene fused to sequence encoding a C-terminal His6-tag. The cells were grown in 2× YT media at 30 °C, and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 16 h at 15 °C. The cells were then lysed by sonication in a lysis buffer (20 mM Na/K-phosphate buffer, pH 7.5, 500 mM potassium acetate, 10 mM imidazole, 0.1% Triton X-100, 1 mM DTT, 1 mM benzimidazole, 100 µM leupeptin, 10 µM pepstatin A and 1 mM PMSF), and was eluted by a gradient of 10–500 mM imidazole in a buffer containing 20 mM Na/K-phosphate buffer, pH 7.5, 300 mM potassium acetate and 5% glycerol. The eluent was further purified using HiTrap Heparin 1 ml (GE healthcare) and CaptoSP ImpRes (GE Healthcare).

#### DNA constructs for single-molecule experiments

**SNR20\* short (–62/+96) promoter sequence for hindering-load assay.** The sequence of the non-template DNA strand was as follows: 5'-GCCGTTTCCGATGGGCCACTCGGTGAAAAACATATAAAAAAGGGCTCTACATTCATTTT TTTTAAATGCCACGAATCTCTTTTCTTTTCGGGTGGATCAAGTGTAGT ATCTGTCTTTTCAGTGTAAACACTGAAATGACCTCAATGAGGCTCATT ACC-3'.

**SNR20\* long (–122/+96) promoter sequence for hindering-load assay.** The SNR20\* long differed from the SNR20\* short by containing an additional 60 bp of DNA between the TATA site and the start site. The sequence of non-template strand of the longer promoter is shown, with the bases underlined not present in the shorter promoter: 5'-GCCGTTTCCGATGGGCCACTCGGTGAAAAACATA TAAAAAGGGCTCTACATTCATTTTTCATCGATGAGTACTTTTACTTGTG TATCAGATTTATTCATTTTGTCTTCTACTTGTTTTTTTTTTAAATGCCACG AATCTCTTTTCTTTTCGGGTGGATCAAGTGTAGTATCTGTCTTTTCAGT GTAACAACTGAAATGACCTCAATGAGGCTCATTACC-3'.

**SNR20\* short (–62/+636) promoter sequence for assisting-load assay.** The SNR20\* short sequence for the assisting-load assay contained an additional 540 bp of DNA downstream derived from the wild-type SNR20 gene relative to the hindering-load assay, as is shown below: 5'-GCCGTTTCCGATGGGCCACT

CGGTGAAAAACATATAAAAAAGGGCTCTACATTCATTTTTTTTAAATGCCC ACGAATCTCTTTTCTTTTCGGGTGGATCAAGTGTAGTATCTGTCTTTTC AGTGTAAACAACTGAAATGACCTCAATGAGGCTCATTACCTTTTAAATTTG TTACAATACACATTTTGGCACCCAAAATAATAAAATGGACGGGAAG AGACTTTTAAAGCAAGTTGTTTTCCGCTAATGTCAGGTCCTCACTACTTT TTGCTGCTATTTTCTTCGCTCATGGTTTCTTCATAAGGCGTTTTATG ATGGTTTTTCGAAATGGTTTTTGAGACGACGGAATCACGAATCTCGG ATCCTTGCTCAAGGTATTGTTTTTGTTCCTCTCGTTGTTGTTTCTATTT TCTTTTTTTAGCTTTCTGTTTCTCCCTTAGTTTGGCTTTTTGCTTCATA CTCTCCCTGTCTTTCCGAGCCGTTTATGTCCAACGCGGGATTTGGTTT TTCTTTATCGATGGGAAGAAATGGTGCTATAGTAGGTTGGGAGATAAT ATTTATGGTATGGGGTGCTAGTGGCGGATGGGGCGCTTTATTGTTGAT TTCTTCGCTCGCTCTCTTTTCTGTTGCGCTGCAAGAGGAAGTTTTC GACTTTGTTATGATTTTTGGTTTGAAGGAAAGGTGTCTTAC-3'.

**Generation of DNA templates.** A 2.7 kb DNA fragment that served as a 'handle' in our single-molecule assay was amplified by PCR from the plasmid pRL702 as previously described<sup>14</sup>, and was subcloned into the pDrive Cloning Vector (Qiagen). To obtain the three DNA constructs used in this study, three different plasmids were constructed, each containing the handle adjacent to one of the following promoter sequences: SNR20\* short (–62/+636), SNR20\* short (–62/+96), and SNR20\* long (–122/+96). For the assisting-load assay, the handle was located upstream of the SNR20\* short (–62/+636) promoter sequence. For the hindering-load assay, the handle was situated downstream of the SNR20\* short (–62/+96) or SNR20\* long (–122/+96) promoter sequence. Regions containing the promoter and the handle were amplified by PCR, using a 5'-digoxigenin labelled primer (IDT) that anneals to the end of the handle, such that PCR products carry a digoxigenin tag on the upstream end of DNA for the assisting-load assay, and on the downstream end of DNA for the hindering-load assay. The generated PCR products were loaded onto TSKgel DEAE-5PW (Tosoh), eluted by a gradient of 0.1–1 M NaCl in a buffer containing 20 mM Tris, pH 7.5, and 2 mM DTT, and concentrated up to 5–10 µM using Vivacon 500 5K MWCO (Vivaproducts), yielding ~0.15–0.3 nmol from 4–8 ml PCR reaction.

**PIC assembly and isolation.** The PIC was isolated as previously published<sup>5</sup> with minor modifications. First, 0.15 nmol of SNR20\* promoter DNA with or without the 2.7 kb handle was separately mixed with 1.5 nmol of TFIIB, 1.5 nmol of TFIIA, 0.8 nmol of TBP, 0.65 nmol of TFIIE, 0.24 nmol of TFIH-ΔTFIIE, and 0.8 nmol of Sub1 in 90 µl of buffer (500) (20 mM HEPES, pH 7.6, 5 mM DTT, 2 mM MgSO<sub>4</sub> and 5% glycerol, with the mM concentration of potassium acetate in parentheses). The mixture was dialysed into buffer (300), buffer (220), buffer (150), and then combined with 0.2 nmol of biotinylated Pol II–TFIIF complex. The mixture was further dialysed into buffer (80), and loaded onto a 10–40% (v/v) glycerol gradient containing 20 mM HEPES, pH 7.6, 5 mM DTT, 2 mM magnesium acetate and 80 mM potassium acetate, and was ultra-centrifuged for 4 h at 48,000 r.p.m. (Beckman SW60 Ti rotor). The presence of the 2.7-kb handle did not significantly affect the efficiency of assembly of the PIC.

**In vitro transcription assay.** Transcription assay was performed as described before<sup>5</sup>. In brief, 1.5 pmol of DNA fragment was combined with 3.7 pmol of TFIIB, 3.7 pmol of TFIIA, 1.5 pmol of TBP, 3.7 pmol of TFIIE, 1.5 pmol of TFIH, 1.5 pmol of Pol II, 2.1 pmol of TFIIF, 2.5 pmol of Sub1 in 5 µl of buffer (300) (50 mM HEPES, pH 7.6, 300 mM potassium acetate, 5 mM DTT and 5% glycerol), diluted with 5 µl of buffer (10) (20 mM HEPES, pH 7.6, 10 mM potassium acetate, 5 mM MgSO<sub>4</sub> and 5 mM DTT), and incubated for more than 1 h on ice. The transcription was initiated by adding an equal volume of buffer containing 20 mM HEPES, pH 7.6, 10 mM potassium acetate, 5 mM MgSO<sub>4</sub>, 10 mM magnesium acetate, 1 U of RNaseOUT, 5 mM DTT, 1.6 mM ATP, 1.6 mM GTP, 1.6 mM CTP, 40 µM UTP and 0.83 µM [ $\alpha$ -<sup>32</sup>P] UTP (2.5 µCi). The reaction was stopped after 15 min by adding 185 µl of stop buffer (300 mM sodium acetate (pH 5.5), 5 mM EDTA, 0.7% SDS, 0.1 mg ml<sup>-1</sup> glycogen, 0.013 mg ml<sup>-1</sup> of proteinase K (Sigma)). Transcripts were precipitated by adding 700 µl of ethanol, dried and analysed by a denaturing 4–12% acrylamide gel.

**Exonuclease footprinting.** Exonuclease footprinting was performed as described before<sup>5</sup>. SNR20\* long (–122/+147) was amplified by PCR in 2 ml reaction using <sup>32</sup>P-labelled upstream primer (5'-GCCGTTTCCGATGGGCCACTC-3') and downstream primer (5'-CCATTTTATTATTTTGGGTGCC-3'), and was purified by electrophoresis in a 2% agarose gel. The labelled DNA (1.5 pmol) was incubated with 3.7 pmol of TFIIB, 3.7 pmol of TFIIA, 2.0 pmol of TBP, 3.7 pmol of TFIIE, 2.0 pmol of TFIH, 2.0 pmol of Pol II–TFIIF complex, and 3.8 pmol of Sub1 in 5 µl of buffer (300) (50 mM HEPES, pH 7.6, 300 mM potassium acetate, 5 mM DTT and 5% glycerol), then combined with 5 µl of buffer (30) (50 mM HEPES, pH 7.6, 5 mM MgSO<sub>4</sub>, 30 mM potassium acetate and 5 mM DTT), and incubated for more than 1 h at 4 °C. The reconstituted PIC was combined with an equal volume of 2× NTP buffer (1.6 mM NTP(s) or 1.6 mM dATP,

50 mM HEPES, pH 7.6, 5 mM MgSO<sub>4</sub>, 30 mM potassium acetate, 5 mM DTT, 10 mM magnesium acetate and 5 U of RNaseOUT) and incubated for 4 min at 30 °C. Exonuclease III digestion was performed with 5–10 U of the exonuclease (NEB) for 9 min at 30 °C, and was stopped by adding 185 µl of stop buffer (300 mM sodium acetate, pH 5.5, 5 mM EDTA, 0.7% SDS, 0.1 mg ml<sup>-1</sup> glycogen, 0.013 mg ml<sup>-1</sup> of proteinase K (Sigma), 0.5 mg ml<sup>-1</sup> salmon sperm DNA (Invitrogen)). DNAs were precipitated by adding 700 µl of ethanol, dried and analysed by a denaturing 6% acrylamide gel.

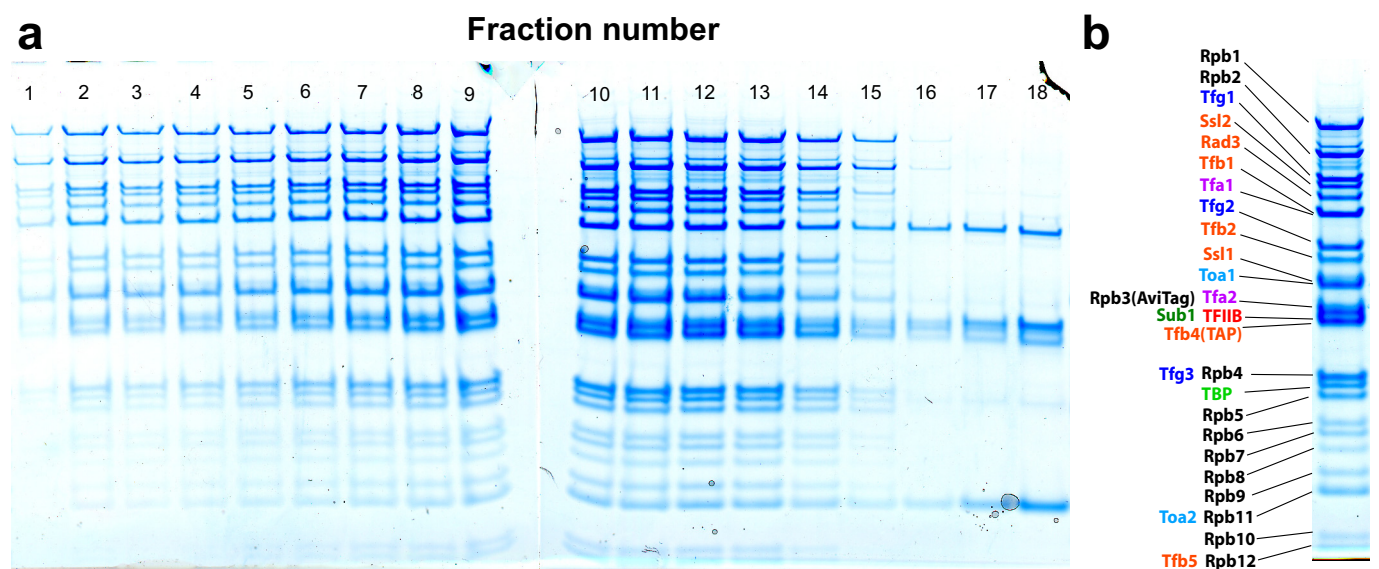
**Data analysis of single-molecule records.** Transcription in the expected direction resulted in a decrease in extension of the DNA tether in the hindering load assay, and an extension increase the assisting-load assay. In both geometries, the change in extension of the DNA tether, which is a function of the applied force, was converted to a distance on the template in bp ( $\sim 0.313 \text{ nm bp}^{-1}$  at  $\sim 4 \text{ pN load}^{14}$ ). The data acquired was smoothed in software by applying a low-pass filter to it (end of pass band = 0.1 Hz; start of reject band = 50 Hz, number of coefficients = 500). To align the records so that motion was defined to start at 0 bp, the mean value of the region  $\sim 1$ –2 s of positional data before the start of processive motion was set to be the starting (0 bp) distance. The velocities of Pol II and TFIIF reported were obtained by dividing the observed change in distance by the time over which the molecule moved. For both Pol II and TFIIF, these velocities were calculated over a region of at least 50 bp that did not contain any resolved pauses ( $>0.1$ – $0.2 \text{ s}$ ). As some of the pauses were short lived, especially owing to TFIIF motion, we did not have sufficient information to characterize the pause lifetimes or distributions, nor could we reliably use previous techniques<sup>14</sup> to get a pause-free velocity. We were also unable to determine pause-free velocities from the distributions of instantaneous velocities<sup>38</sup>, as there was often insufficient data to obtain reliable fits. For the TFIIF records, velocities of the scanning state and fast states, when observed, were occasionally calculated by examining different regions of the molecules that were separated in time by a relatively sharp transition (change within 0.5 s) in velocity. To estimate the processivity of TFIIF during scanning, we only included

molecules that travelled at least 30 bp, sufficient to move beyond the noise threshold and extend past the distance of the open complex ( $24 \pm 2 \text{ bp}$ ).

**Calculating expected open-complex distance.** Biochemical studies have shown that the minimal distance of a TSS from the TATA box is about 30 bp, and that transcription begins at this location in the initial open complex<sup>4</sup>. In the structure of the closed PIC<sup>29</sup>, this location in the promoter DNA is about 80 Å from the nucleotide addition site of Pol II. Open complex formation must bring the TSS to the nucleotide addition site, which therefore requires drawing 80 Å of downstream DNA into the Pol II cleft. In the structure of a transcribing complex<sup>20</sup>, all but 3 bp of downstream DNA are double-stranded, and 80 Å corresponds to 24 bp of dsDNA, the same as the initial movement observed in the hindering-load assay.

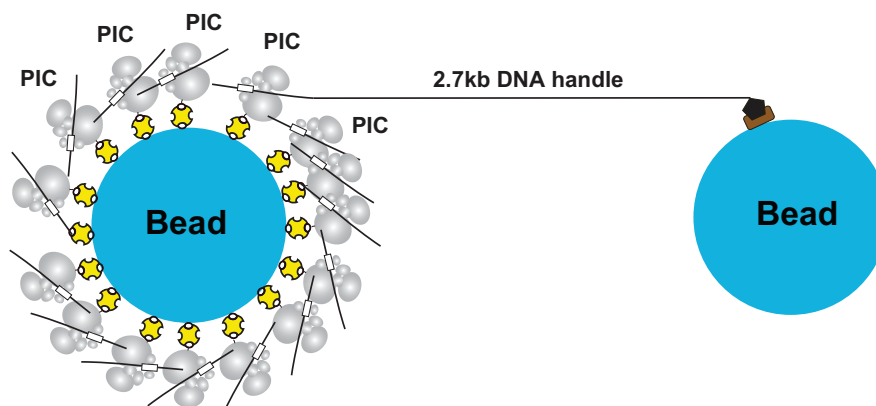
31. Laybourn, P. J. & Dahmus, M. E. Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* **265**, 13165–13173 (1990).
32. Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. D. CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* **67**, 1223–1230 (1991).
33. Koslover, D. J., Fazal, F. M., Mooney, R. A., Landick, R. & Block, S. M. Binding and translocation of termination factor rho studied at the single-molecule level. *J. Mol. Biol.* **423**, 664–676 (2012).
34. Murakami, K. *et al.* Tfb6, a previously unidentified subunit of the general transcription factor TFIIF, facilitates dissociation of Ssl2 helicase after transcription initiation. *Proc. Natl Acad. Sci. USA* **109**, 4816–4821 (2012).
35. Ge, H. & Roeder, R. G. Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* **78**, 513–523 (1994).
36. Henry, N. L., Bushnell, D. A. & Kornberg, R. D. A yeast transcriptional stimulatory protein similar to human PC4. *J. Biol. Chem.* **271**, 21842–21847 (1996).
37. Malik, S., Guermah, M. & Roeder, R. G. A dynamic model for PC4 coactivator function in RNA polymerase II transcription. *Proc. Natl Acad. Sci. USA* **95**, 2192–2197 (1998).
38. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437–447 (2003).





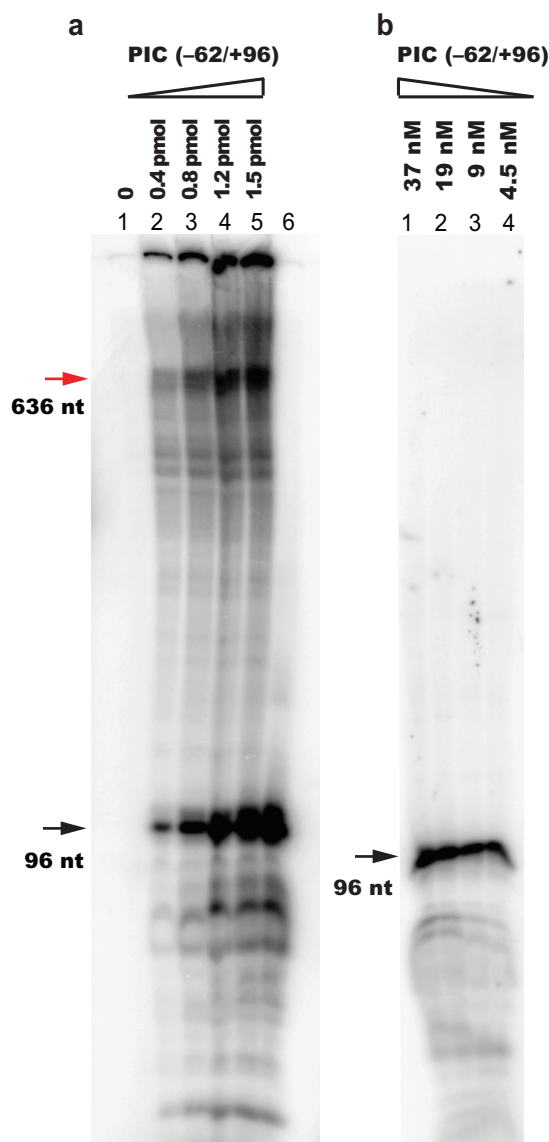
**Extended Data Figure 1 | The 29-component PIC assembled on *SNR20\** short promoter.** **a**, PIC excluding the kinase domain (TFIIK) was assembled on *SNR20\** short (adjacent to the 2.7-kb downstream handle sequence) and sedimented on a glycerol gradient; fractions were analysed by SDS-PAGE. **b**, The results from fraction 12, annotated in detail, indicate that all PIC

components were retained, confirming that the complex reconstituted fully from the component proteins. The subunit(s) of Pol II are labelled in black, TFIIF in blue, TFIIE in magenta, TFIH in orange, TFIIA in cyan, TFIIB in red, TBP in light green, and Sub1 in dark green. TFIIK (3-subunits) was later added to the PIC.

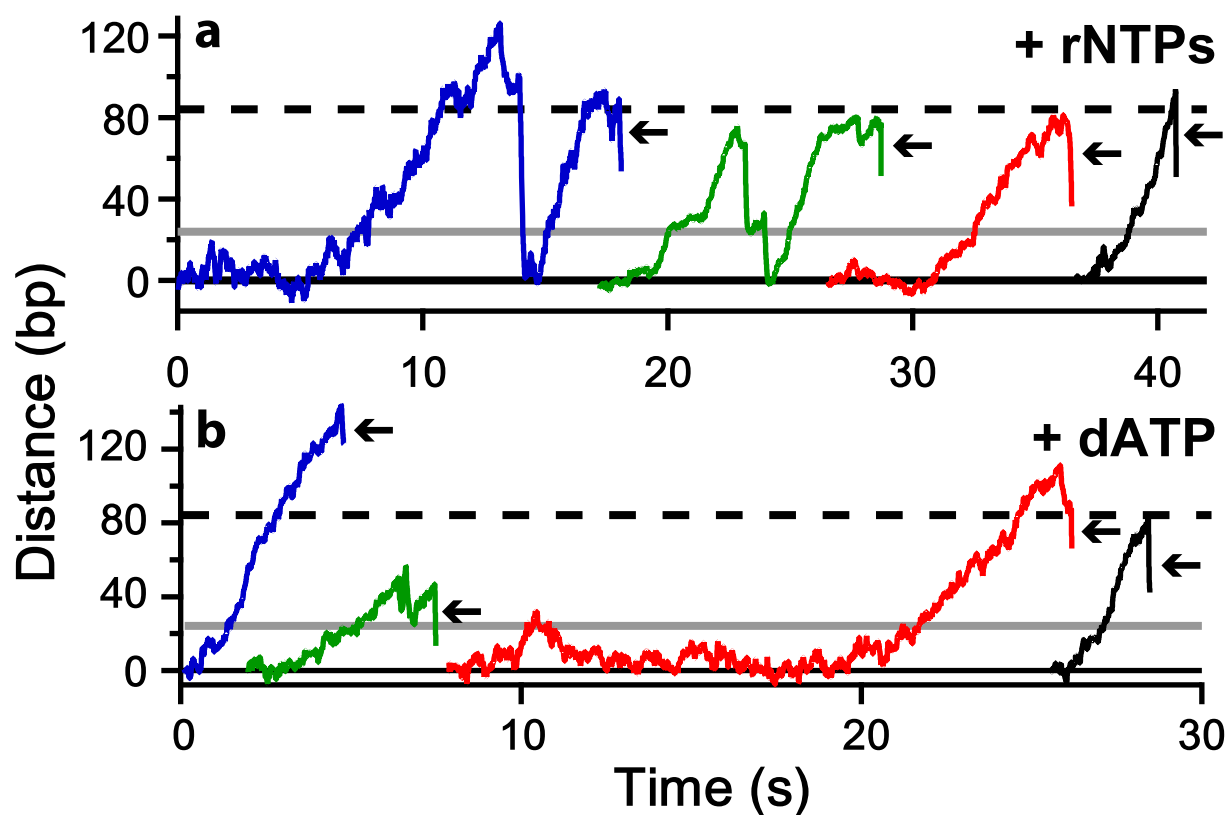


**Extended Data Figure 2 | Schematic diagram showing assembly of dumbbells, in cross-section.** PICs were attached to one bead via biotin-avidin linkages (yellow). To form dumbbell tethers, the other end of a small fraction of the PICs (4%) had digoxigenin linkages that could be tethered to anti-

digoxigenin-coated beads (black and brown) via a 2.7-kb DNA handle. PICs not involved in tether formation served to increase the local concentration of PIC components.



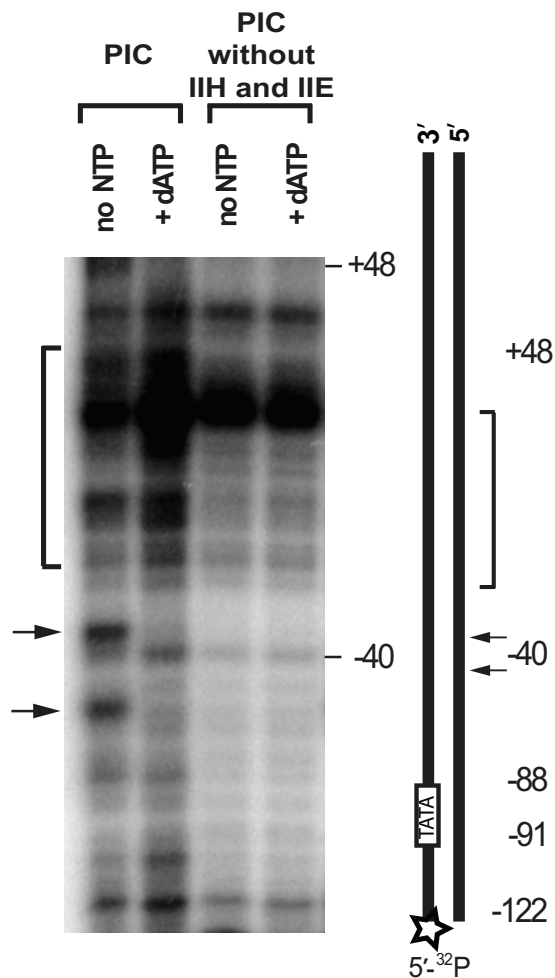
**Extended Data Figure 3 | Run-off transcription under single-molecule assay conditions.** **a**, Isolated PICs (0.1 pmol), formed on the *SNR20\** short promoter fragment fused to the transcription template (covering the region -62/+636), and attached to a 2.7 kb DNA handle, was combined with increasing amounts of PICs assembled on the *SNR20\** short promoter, but without the handle, hereafter referred to as PIC (-62/+96). These constituents were incubated with an equal volume of a 2X NTP solution (10 ml) containing 1.6 mM ATP, 1.6 mM GTP, 1.6 mM CTP, 40 mM UTP, and 0.83 mM [ $\alpha$ - $^{32}$ P] UTP (2.5  $\mu$ Ci). The resulting transcripts were analyzed by gel electrophoresis. PICs fused to the DNA handle failed to support transcription alone (lane 1), but transcription activity was restored (red arrow) when a 4-fold (lane 2), 8-fold (lane 3), 12-fold (lane 4), or 15-fold (lane 5) excess of PIC (-62/+96) was added to the reactions. In lane 6, the reaction contains 1.5 pmol PIC (-62/+96). The 96-nt run-off transcription from PIC (-62/+96) is indicated (black arrow). A 25-fold excess of PIC (-62/+96) was used for single-molecule assays (Extended Figure. 2). **b**, 1.5 pmol aliquots of PIC (-62/+96) were introduced into different volumes of transcription buffer, such that assayed concentration of PIC varied from 37 nM to 4.5 nM. Transcription efficiency (run-off band, black) decreased with PIC concentration from ~18% to just 2–3%. The low concentrations used in single-molecule assays (<1 nM) could not be assayed directly using gels, but we expect that the transcription efficiency is correspondingly low.



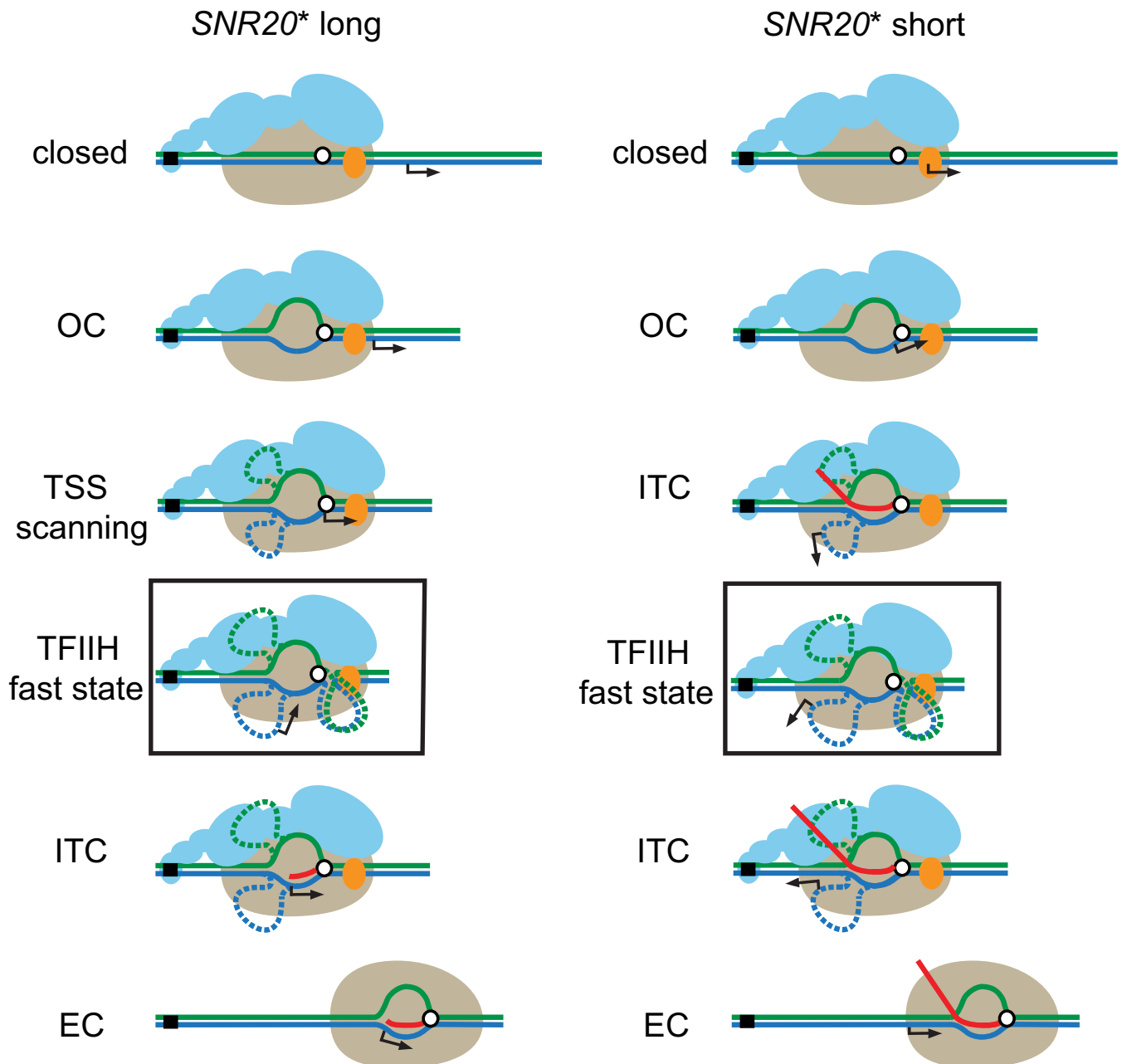
**Extended Data Figure 4 | Records of TFIID scanning on SNR20\* long with rNTPs or dATP.** **a, b,** Just as for SNR20\* short (Fig. 3), the longer promoter shows TFIID scanning with both rNTPs (**a**) or dATP only (**b**), after which

either the PIC dissociates (black arrows), or the bubble collapses to the closed (blue and green records) or open (grey line) complex and TFIID moves again. The dashed line indicates the position of the TSS (+1).





**Extended Data Figure 5 | Exonuclease III footprinting assay of the PIC on *SNR20\** long.** In the absence of nucleotides *in vitro*, PIC complexes bound to the *SNR20\** long promoter produced barriers to exonuclease III digestion located ~50 bp downstream of the TATA box (about -40 nucleotides from the TSS, black arrows). These barriers depended on the presence of TFIID and also TFIIE, which interacts with TFIID. After the addition of dATP, the barriers disappeared, and the bands at pause positions were intensified between positions -30 and +30 (~60–120 bp downstream of the TATA box, bracket).



**Extended Data Figure 6 | The transcription initiation pathway for *SNR20\** long (left) and *SNR20\** short (right) promoters.** Left, a model for the initiation pathway on the *SNR20\** long promoter. States starting from the top: Pol II (beige) with attached GTFs (blue) and Ssl2 (orange) binds in its 'closed' form to the promoter element upstream of the TSS (arrow) on the DNA template (green and blue lines). Positions of the enzyme active site (open white circle) and TATA box (closed black square) are indicated. Unwinding by TFIIH produces an open complex (OC) that leads to bubble formation. Arrival of the open complex at the TSS owing to scanning, driven by TFIIH, leads to the formation of an extended bubble (dashed lines indicate the speculative position of single-stranded DNA). If the complex fails to recognize

the TSS, it can be driven beyond it by TFIIH, resulting in a 'fast state' that produces no RNA but advances at roughly twice the normal rate (black box; see text). When Pol II recognizes the TSS, it begins transcription of RNA (red line), corresponding to the ITC. Formation of the ITC leads to bubble collapse, followed by the loss of GTFs and transition to the elongation complex (EC). Right, corresponding model for the initiation pathway on the *SNR20\** short promoter. Similar states as for *SNR20\** long. In this case, the open complex does not need to scan for the TSS, which is found within its DNA footprint. As a consequence, the ITC can form and begin RNA synthesis once the active site has recognized the TSS. A longer segment of RNA can thereby be produced before the transition to the elongation complex (EC).

CORRIGENDUM

doi:10.1038/nature14609

Corrigendum: Passenger deletions generate therapeutic vulnerabilities in cancer

Florian Muller, Simona Colla, Elisa Aquilanti, Veronica E. Manzo, Giannicola Genovese, Jaclyn Lee, Daniel Eisenson, Rujuta Narurkar, Pingna Deng, Luigi Nezi, Michelle Lee, Baoli Hu, Jian Hu, Ergun Sahin, Derrick Ong, Eliot Fletcher-Sananikone, Dennis Ho, Lawrence Kwong, Cameron Brennan, Y. Alan Wang, Lynda Chin & Ronald A. DePinho

Nature 488, 337–342 (2012); doi:10.1038/nature11331

In this Article, during the preparation of Figures 2d and 3a, we processed digital western blot scans to remove duplicate or otherwise irrelevant lanes from single-blot images. Although all excisions/mergers originated from the same gel, these figure constructions should have been explicitly pointed out. Here we present the unprocessed scans (Supplementary Information) and amended figures (Figs 1 and 2). Figure 1 of this

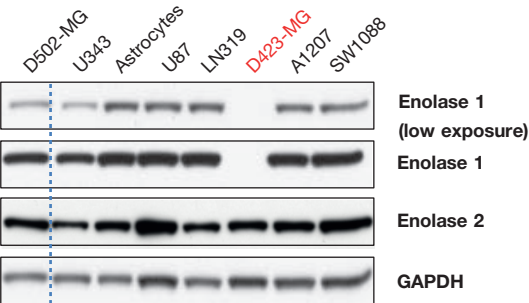


Figure 1 | This is the corrected Fig. 2d of the original Article, with excision indicated.

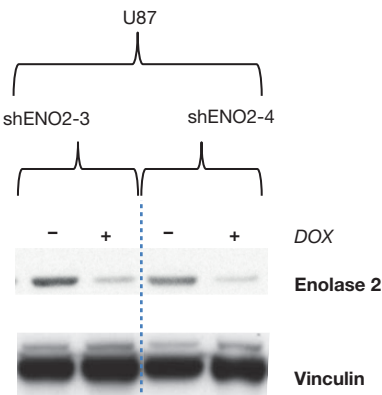


Figure 2 | This is the corrected Fig. 3a of the original Article, with excision indicated.

Corrigendum shows the corrected Fig. 2d, in which a duplicate run of cell line D423-MG (lane 2 in the original) was excised between cell lines D502-MG and U343 (lanes 1 and 3 in the original) and the ensuing halves of the blot were spliced together (lanes 1 and 3 in the original blot). This is now indicated by a dashed line. Similarly, Fig. 2 of this Corrigendum shows the corrected Fig. 3a, in which for the cell line U87, an additional non-targeting short hairpin RNA control (original lanes 7 and 8) was excised with the remaining halves of the blot and merged, which is now indicated by a dashed line. We also note that in the published Fig. 3a, lanes 1 and 2 of the original U87 vinculin blot were accidentally used as the loading control for shENO2-4 (lanes 9 and 10 of the original unprocessed ENO2 blot in the Supplementary Information), and lanes 9 and 10 of the vinculin blot should have been used as the correct loading control lanes. The correct loading control lanes are now shown (Supplementary Information). None of these corrections alter the original meaning of the experiments, their results, their interpretation, nor the conclusions of the paper. We apologize for any confusion this may have caused to the readers of Nature.

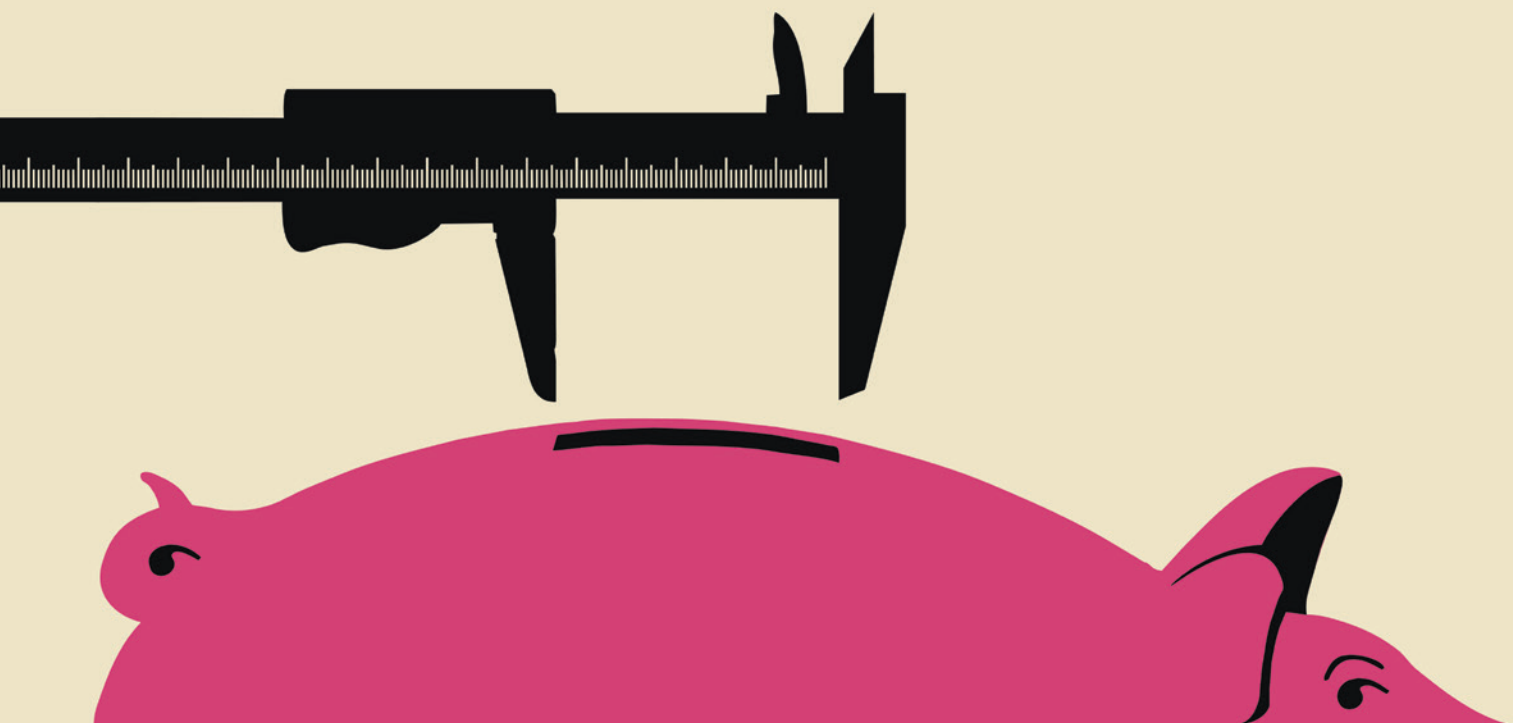
Supplementary Information is available in the online version of this corrigendum.

# CAREERS

**TILT THE ODDS** A physicist uses his skills in statistical analysis in finance **p.281**

**SELF PROMOTION** Social media is a powerful tool for promoting work [go.nature.com/rsejgi](http://go.nature.com/rsejgi)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



## EMPLOYEE BENEFITS

# Plight of the postdoc

*As institutions attempt to redefine the postdoctoral position, early-career researchers are joining together to wage a battle for proper benefits.*

BY HELEN SHEN

**A**nna Kalashnikova is a master of efficiency. The University of California (UC), Davis, postdoc arrives at her laboratory by 8 a.m., knowing that she must finish her experiments by 5:35 p.m. every night. After a long day spent studying how disease-related modifications of DNA-packaging ‘histone’ proteins are regulated, she must cycle 25 minutes to her son Maxim’s child-care facility, which closes at 6 p.m. sharp.

For every minute that she is late to pick him up, Kalashnikova must pay extra fees. And as a single mother supporting her child on a postdoc’s salary, there is little wiggle room. Nearly half of her monthly income goes towards child

care, and one-third covers rent and utilities at the house she shares with a roommate. “There’s this constant stress, because if something unexpected happens, we’d be in big trouble,” she says.

Kalashnikova is one of many postdocs in the UC system who are hoping that their circumstances may soon improve. They are in the midst of negotiating a new contract with the university administration — with the current contract set to expire on 30 September, the UC postdocs’ labour union is pushing for improvements on many fronts, including salary, career development and child-care support.

The negotiations come at a time of great turmoil for postdoctoral researchers worldwide, as academic science faces a critical oversupply of postdocs and a shortage of tenured faculty

positions. And as early-career researchers find themselves stuck with low pay and minimal benefits for longer periods, postdocs and their advocates at several institutions, including the University of Maryland (UMD) in College Park and the Howard Hughes Medical Institute (HHMI) in nearby Chevy Chase, are fighting — with varying degrees of success — for greater benefits, and standardized titles and rights (see ‘A postdoc by any other name’).

“These are not new problems, but they are perhaps being more acutely felt now. As grant paylines remain low, universities are hiring fewer tenure-track faculty and postdocs remain an abundant source of low-cost labour,” says Keith Micoli, chairman of the board of the US National Postdoctoral Association in ►

NEIL WEBB/GETTY



► Washington DC, a non-profit organization that has been advocating for postdocs' rights since 2003 (see *Nature* <http://doi.org/65x>; 2012). "The situations at the Howard Hughes Medical Institute, University of Maryland and the University of California system are the latest examples of challenges facing postdocs," he says.

## BATTLE FOR BENEFITS

At institutions across the globe, lab heads have had to support a ballooning population of postdocs with tight, government-funded research budgets. Even some major, privately funded institutions have recently cut back on benefits for postdocs. In September 2014, for example, the HHMI angered many postdocs when it announced that it would reduce some of their long-term benefits beginning in 2015.

In an e-mail to employees, the institute explained that postdocs spend a short time at the HHMI and so their benefits priorities often differ from those of other employees.

At the HHMI's Janelia Research Campus in Ashburn, Virginia, postdocs bristled at the institute's reasoning. "They're operating under the assumption that the postdoc is a short, transient position, when the truth is that most postdocs are lasting about five years," says neuroscientist Eric Yttri, co-president of the Janelia Association of Research Scientists, which represents postdocs and other staff scientists.

As of this year, postdocs stopped receiving retirement contributions from the HHMI — a standard benefit given to most other employees that equates to 5% of their annual salaries. In addition, employees hired after the start of 2015 no longer receive 'benefits credits', an income supplement that HHMI gave twice a

month in the past to help offset health-care costs. The credits will continue to accrue for employees who were hired before 1 January 2015, but will remain frozen at 2014 levels.

In exchange for these cuts in long-term benefits, HHMI revamped its employee health-insurance programme, introducing new options that were intended to be more affordable for postdocs. Nevertheless, some postdocs saw their health-care costs rise in 2015 compared to 2014. Cory Schreckengost, director of administrative operations at Janelia Farm, declined to disclose specific details about the organization's employee benefits and costs, but emphasized in an e-mail that "HHMI took mitigating steps specifically for postdocs to limit the effects of the rising medical care costs that impact all HHMI employees".

Janelia postdocs raised vocal objections to this raft of changes, which in turn brought about some concessions. For example, the HHMI added a US\$1,500 annual supplement for postdocs, to be used at their discretion. But the institute remained steadfast on cutting the retirement benefit, citing the same practice elsewhere. "Most universities and research institutes do not contribute to retirement accounts for postdoctoral associates, and HHMI has chosen to move to this standard," the institute wrote in an e-mail to employees.

Despite signs of a cool climate for postdocs overall, UC postdocs have prevailed over economic challenges in the past. They formed their union in 2008 and secured their first

contract with the university in 2010 — all in the middle of a state budget crisis.

Among other advances, the initial five-year agreement provided salary increases, retirement contributions and guaranteed time off for holidays and for personal or medical reasons. "Of course, there is a lot of room for improvement," says Anke Schennink, president of the United Auto Workers local affiliate that represents the UC system's roughly 6,000 postdocs.

The union has already made some progress in its second round, reaching a preliminary agreement with the university on 6 August that would secure postdocs' right to pursue career counselling and career-development activities on paid time.

Given the scarcity of tenure-track positions, the next step for many postdocs will probably involve an exit from academia. "Statistically, it's not in any individual postdoc's favour to be completely focused on any one career path. Having this time put onto paper would basically recognize that a postdoc is a training position, and that it's important for us to work on career development," says Jessica Lao, a postdoc at the University of California, San Francisco (UCSF), and a member of P(ostdoc)-Value\*, a grassroots postdoc-advocacy group at UCSF. Lao leads the group's efforts to promote UCSF's career services and pilot programmes for postdocs to tour biotechnology companies and gain hands-on experience in other non-academic careers.

## FAMILY MATTERS

Child care is another priority. Many graduate students across the ten campuses of the university system already qualify for financial support for child-care expenses — up to US\$900 per quarter or \$1,350 each four-month semester. But postdocs do not have such benefits.

Also on the bargaining table is salary, a perennial hot-button topic. For UC postdocs, as for many of their US peers, minimum salaries are pegged to guidelines published by the US National Institutes of Health. But stipends that are equivalent in value can mean vastly different standards of living, depending on where the postdoc resides. "California is expensive, if you compare it in terms of cost of living around the country," says Schennink. "We think postdocs should receive fair compensation," she says.

The high cost of living in California has proved particularly challenging for Abby Kroken, a postdoc at the University of California, Berkeley, who now spends more than 60% of her take-home pay each month on housing, compared with 30% when she was a graduate student at the Medical College of Wisconsin in Milwaukee.

Kroken had carefully studied housing rates and living expenses before coming to UC Berkeley in January 2014 to study bacterial eye infections. And although she thought she was prepared, Kroken could not predict



Molecular biologist Anna Kalashnikova takes time away from the lab to enjoy the outdoors with her son.

MICHAEL VOGEL

## IDENTITY ISSUES

*A postdoc by any other name*

What is a 'postdoc'? In academia, the lack of a standard definition can create a host of problems for researchers.

Some postdocs are classified as university employees, eligible for standard benefits such as health care, child-care support and retirement contributions. But many fall into a hodge-podge of trainee or temporary-worker categories that do not qualify for all the benefits enjoyed by graduate students, faculty members and staff. Often, these ill-defined postdocs lack administrative offices that are dedicated to their professional development, fair treatment and job security.

But standardizing the postdoc position is no simple feat — especially when it means extra costs for the institutions and the individual investigators that employ postdocs. Earlier this year, the University of Maryland (UMD) administration in College Park ran afoul of its life-sciences professors when it tried to eliminate one of two hiring categories for postdocs. The now-defunct category was a contract position with a faculty title and few benefits, which made it a less-expensive option for principal investigators. Only 15% of UMD postdocs fell into this class, but it had commonly been used to hire biomedical postdocs.

"Lots of life-sciences faculty were responding to National Institutes of Health and National Science Foundation budget cuts. They were pinching pennies," says Jonathan Dinman, chair of cell biology and molecular genetics at UMD.

The other category, a non-tenured faculty position, provided postdocs with standard health and retirement benefits, paid medical leave and tuition remission for employees and family members — at a cost that many biomedical lab heads deemed untenable. In a letter to the

university president, more than 130 life scientists wrote that forcing them to use the latter category would add expenses that could not be justified to governmental granting agencies, and would lead to personnel cuts and decreased productivity — amounting to a "death spiral".

But others, including UMD astronomer Marc Pound, argued in favour of the benefits expansion. "Postdocs are kind of a silent majority on campus. They come here for maybe three or six years and move on, and they never really have advocates amongst themselves," says the senior research scientist.

Ultimately, the administration created a new classification scheme that started on 1 July — one that guarantees all postdocs some benefits, but enables lab heads to offer a smaller starting benefits package for less-experienced postdocs. 'Post-doctoral associates' will receive the complete benefits package previously offered to most postdocs. 'Post-doctoral scholars' will get the same benefits, except for tuition remission — which had tended to be the most expensive and unpredictable expense for lab supervisors.

Postdocs can be hired directly into the associates category, but lab heads can also choose to hire early-career postdocs at the scholars level. After three years, however, those postdocs must be promoted to associates if their supervisors wish to renew their contracts. And after a total of six years in either category, postdocs must advance to a research-scientist track.

"It's still a mandate, but now we've got three years to adjust and figure out how to do it," says Dinman, a signatory of the letter. "It has increased the cost of doing business; that's for sure, but in the end, I think the right thing was done." **H.S.**

that her husband, who relocated with her, would be unable to find work for about 11 months. Between the high cost of living in Berkeley and the couple's student-loan obligations, they soon depleted their savings and had to borrow money from their parents to make ends meet.

"It felt like I'd made a gigantic financial mistake in trying to advance my career," says Kroken. For the first time, after focusing exclusively on an academic-research career, she began to consider a job in industry. She also thought about moving back to Wisconsin.

But things began to turn around last December, when Kroken's husband found work as a technical writer. By following

a strict budget, the couple is now close to restoring their previous savings.

Kroken says that her supervisor has given her much-needed encouragement to continue pursuing an academic career. But despite her improved outlook, she says that the past year and a half has underscored just how important adequate compensation is for her continued professional development. "I do want to be a professor, I do like research and I even like writing grants," says Kroken. "I don't want to have to leave this career path because I can't afford to do it." ■

**Helen Shen** is a freelance writer in Sunnyvale, California.

## TRADE TALK

### Fund manager



**Ben Peters** is an investment director at Evenlode Income, an independent fund-management company in Chipping Norton, UK. He explains how his PhD in physics helped

to smooth his transition into the world of finance and investments.

#### Why did you leave academia?

I enjoyed my PhD programme in nanophysics and wouldn't have been averse to staying in academia. But having to reapply for funding every few years didn't much appeal to me, and halfway through my programme, I became interested in investment management through my brother-in-law, Hugh, whom I now work with. So after graduating in 2008, I moved into this industry.

#### How has your PhD work helped you in your role as a fund manager?

Mathematical and statistical-analysis skills are highly valued and important in this industry. They helped me to get through the door. I constructed a method for quantitative analysis of companies' financial information, and I use statistical techniques to look at the risk in any investment. I also developed a lot of soft skills by doing research, particularly on collaborative projects. If you're an experimental scientist, as I was, you have to be flexible — you might have to change research directions and all PhD students learn how to organize themselves and react to what is going on — it's a good skill for this field.

#### What do you enjoy about your job?

I'm finding out how the world works. As a physicist, I was beginning to understand the material world, but this is more about the human and economic worlds. The interactions of people around the globe and what they do and how they create value — I find that fascinating. As a fund manager, I have to figure out how the world works and how it might evolve over time, but also accept the extremely uncertain nature of economic systems. So I have had to develop an investment process that ultimately results in action — making an investment — while knowing that it is a game of tilting the odds, rather than one of certainties. ■

#### INTERVIEW BY JULIE GOULD

This interview has been edited for length and clarity.



# NEURAUGMENT, VERB

*Standard United Galactic Treaty Dictionary, 2nd edn, Caelum Univ. Press (11 June 2287).*

BY FELICIA DAVIN

**neuraugment, v.**

Full entry. All quotations shown.

**Pronunciation:** Standard Treaty ['nu:ʌɡmənt], N Earth Eng ['nɜ:ʌɡmənt]

**Forms:** neuraug, neurog

**Etymology:** <English *neural augmentation* any technological enhancement to the human brain (21st cent.), <ancient Greek νῆρρον nerve (see neuro- comb. form) + -al suffix and <Old French *augmentacion*, see AUGMENT, v. Compare *neuraugmentation* n., *neuraugmented* or *neurogged* adj.

1. *trans.* To enhance the function of the brain through biomechanical implants of any type.

2098 *Newz.ly* 3 January: Dr Sharma proposes to, in her words, “neuraugment” the civilian public. She acknowledges that previous attempts to do so have gone awry, but insists that her method is safer and less obtrusive than any of her predecessors. “Initially, I wanted to help people who were unable to retain information,” she says. “But this could be so much more. It could help us learn each others’ languages.”

2108 G. RODRIGUEZ *Parenting in the New Century*: It’s all the rage among the rich and famous to neuraugment themselves and even their young children. Youngsters with perfect pitch and eidetic memories are common in the climate-controlled enclaves of the Pacific Northwest. I once met a six-year-old who had read *Madame Bovary* in the original French. Her review? “Boring.” Oh, to be six and feel such crushing *ennui*.

2132 L. JOHNSON *North American Educators’ Quarterly* (Dry season): Many teachers who can’t afford to neuraugment themselves have trouble connecting with a generation of students for whom it is the norm, and there has been much discussion of the future of public education.

2147 *Shermer Morning News* 30 April: Ms Alcantara has refused to neuraugment herself, saying: “I don’t trust it. How do you know what they’re really putting in you? I don’t need anybody else in my head.” She is now heading the Natural Brain movement.

2. *trans.* To join with one or more individuals through a type of neuraugmentation allowing

instant mental communication, see NEURALINK, n.

2120 *seoul.times* 15 March: Dr Park and his collaborator Dr Leary propose to improve upon Dr Sharma’s methods. “It’s great, what she did,” Park says. “Incredible, really. But we can do so much more. Why stop at neuraugmenting one brain?” Leary picks up right where Park leaves off: “Why not neuraugment two brains, or more than two? Think of the computational power we could access through our neuralink.”

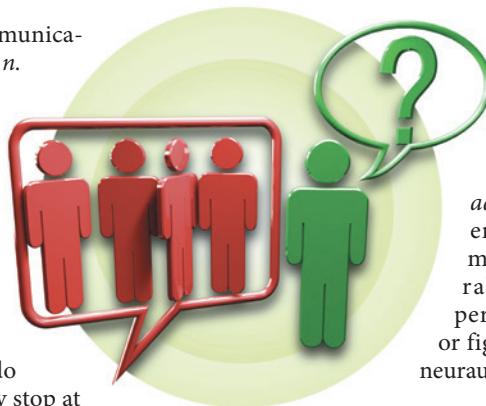
2158 *politics/anon* 9 September: The United Galactic Treaty — a grand name for what’s barely more than a dozen Earth nations and a space station that got a wild hair to design its own flag — well, anyway, the Treaty’s going to recruit citizens by offering to neuraugment them for free if they’ll join up. The whole thing’s doomed if you ask me, as they can only link three or maybe four people at a time without one of them crashing.

2163 *Galactic Treaty Daily* 2 July: Tragedy has struck aboard the *Jespersen*, where five young recruits who agreed to neuraugment have suffered aneurysms.

2169 *Galactic Treaty Daily* 28 December: “It is completely safe to neuraug two people,” the doctor insists. “Some people have the constitution — we don’t know why just yet — that allows them to neuraug three people. I am confident that more is possible, but the safety of our patients is always our first concern.”

2225 A. CHEN *Sex and Love in the Neuraugmented Age*: Georgia, who has asked that I change her name to protect her privacy because her parents do not approve of her lifestyle, neuraugmented four other people, and she loves them all. “I couldn’t live without them,” she tells me. “It sounds so lonely.”

2284 L. MARTINEZ *Collected Correspondence*: He left me a diamond ring on the kitchen counter, attached to a note that said: “I read people used to give these to each other back in the days of civil marriage licences and reading vows and everything



— thought it was cute. Neurog me?”

## DERIVATIVES

**neuraugmented,**

*freq. neurogged*, *adj.* Possessing brain enhancements; intimate with or inseparable from another person, either literally or figuratively linked via neuraugmentation.

2150 ANONYMOUS *Yakt* 2 August: All these neuraugmented jerks think they’re better than us.

2237 R. RAI *Stars Among the Stars* 12 February: “I hate these parties,” the actress sighs, taking a sip of champagne imported from the surface. The party is silent except for the clink of glasses. We are the only two people speaking aloud. “Everyone’s so neurogged and full of themselves. Or full of each other, I guess. I just want something real, you know?”

2278 *Galactic Treaty Daily* 17 May: The staff members agree among themselves that the Secretary General and the Chair of Military Operations are so neurogged that approaching one of them is essentially the same as approaching both of them, but this was not always the case.

**neuraugmentation, n.** Any technological enhancement to the brain.

2136 K. NSONWAH *A Natural and Technological History of the Brain*, epilogue: Dr Sharma has been reclusive in her retirement, but she graciously invited me into her home to speak with me. She offered me tea in her tidy succulent garden, and she was so welcoming that I could not resist asking a personal question. “Is it true you’ve never had any neuraugmentation?” I said. Dr Sharma shrugged. “I never felt the need. Aisha has a perfect memory,” she said, speaking of her partner of decades. “So if I forget something, I just ask her. As for the neuralink, oh, I don’t know.” She paused to adjust her shawl. “It’s nice, don’t you think? Talking?” ■

**Felicia Davin** is a linguist and translator in western Massachusetts. Her fiction has previously been published in *Lightspeed*. You can find her on twitter @FeliciaDavin.

ILLUSTRATION BY JACEY